

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Principles and Advances in Analysis of Ribonucleic Acid Sequence Using Nanopores

Permalink

<https://escholarship.org/uc/item/24b97646>

Author

Smith, Andrew Martin

Publication Date

2017

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**PRINCIPLES AND ADVANCES IN ANALYSIS OF RIBONUCLEIC
ACID SEQUENCE USING NANOPORES**

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Andrew M. Smith

June 2017

The Dissertation of Andrew M. Smith is
approved:

Professor Glenn Millhauser, Chair

Professor Mark Akeson, Research Advisor

Professor Michael Stone

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by
Andrew M. Smith
2017

Table of Contents

List of Figures	v
List of Tables	vii
Abstract	viii
Dedication	x
Acknowledgements	xi
Introduction	1
<i>Background on nanopore sequencing</i>	1
<i>Background on RNA sequencing</i>	4
<i>Outline of the Dissertation</i>	9
<i>Individual Contributions</i>	10
Capture, unfolding and detection of tRNA using a nanopore device	12
<i>Introduction</i>	12
<i>Results</i>	15
<i>Discussion</i>	31
<i>Methods</i>	36
<i>Acknowledgements</i>	41
Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing	42
<i>Introduction</i>	42
<i>Results</i>	44
<i>Conclusion</i>	59
<i>Methods</i>	59
<i>Acknowledgements</i>	66

Active translocation of SV40 Large T-antigen helicase with DNA and RNA substrates atop α -hemolysin nanopore	67
<i>Introduction</i>	67
<i>Results</i>	70
<i>Conclusion</i>	84
<i>Methods</i>	85
<i>Acknowledgements</i>	88
Bibliography	89
Protocol for selective preparation of RNA for nanopore sequencing	102
Protocol for preparation of LTag 262-627 helicase fragment	110

List of Figures

Figure 1.1: Some modifications commonly found in RNA	6
Figure 2.1: Strategy for constructing adapter-linked tRNA for nanopore analysis	17
Figure 2.2: Adapted tRNA-dependent ionic current blockades observed during single-channel α HL nanopore experiments.	19
Figure 2.3: Nanopore capture of adapted RNA complexed with non-catalytic ϕ 29 DNAP	21
Figure 2.4: End-to-end translocation of adapted RNA(hairpin) constructs through α -HL nanopore.	23
Figure 2.5: Adapted tRNA ^{fMet} and tRNA ^{Lys} translocate through the α -HL nanopore.	29
Figure 2.6: Classification of tRNA molecules using duration and mean ionic for regions I-III of adapted tRNA translocation events.	32
Figure 2.7: Examples of ionic current traces during translocation of RNA hairpins translocation through α -HL with the dual abasic adapter, the 5' mono-abasic adapter or the 3' mono-abasic adapter.	34
Figure 2.8: Examples of adapted tRNA ^{Lys} and tRNA ^{fMet} ionic current traces during ϕ 29 DNAP-regulated translocation through α -HL nanopore.	35
Figure 3.1: Design and testing of an oligonucleotide adapter to prepare 16S rRNA for nanopore direct RNA strand sequencing	45
Figure 3.2. Direct nanopore sequencing of individual <i>E. coli</i> 16S ribosomal RNA strands.	46
Figure 3.3: Alignment metrics for Enolase 2 polyA calibration strand and <i>E. coli</i> 16S rRNA.	49
Figure 3.4: Detection of m7G modifications in <i>E. coli</i> 16S rRNA.	52
Figure 3.5: Confirmation of m7G at positions 527 and 1405 in <i>E. coli</i> 16S rRNA.	53
Figure 3.6: Inference of pseudouridine in <i>E. coli</i> 16S rRNA direct sequencing reads	55

Figure 3.7: Direct 16S rRNA sequencing discriminates among microbes and can detect <i>E. coli</i> 16S rRNA at low concentration in a human RNA background.	58
Figure 4.1: Structure of Simian Virus 40 (SV40) Large T-antigen (LTag) helicase fragment	68
Figure 4.2: Formation of activity helicase complexes by construct LTag 262-627 under buffer conditions compatible with nanopore analysis	71
Figure 4.3: Putative controlled translocation of a DNA hairpin through α -hemolysin (aHL) nanopore in the presence of active LTag helicase fragment	73
Figure 4.4: Putative controlled translocation of RNA hairpins through α -hemolysin nanopore in the presence of active LTag helicase fragment	76
Figure 4.5: Amplitudes associated with single-nucleotide displacement of hairpin substrates can be identified from RNA induced-pausing	80
Figure 4.6: Reproducible ionic current amplitudes are associated with single-nucleotide displacement of DNA and RNA hairpin substrates during active LTag-mediated translocation through aHL nanopore	81
Figure 4.7: Approximate DNA translocation rates by LTag helicase fragment atop aHL nanopore measured over 13nt window defined by two high amplitude markers	83

List of Tables

Table 2.1: RNA(hairpin) nanopore translocation events classified by detection of leading and trailing high current markers.	26
Table 3.1: Error rate profile for Enolase 2 transcript and <i>E. coli</i> 16S rRNA	48
Table 3.2: Detection of nucleotide variants in MinION <i>E. coli</i> 16S rRNA sequencing data using marginCaller	50
Table 4.1: Quantification of active LTag/hairpin complexes captured on a nanopore before the 5-abasic block is translocated through the pore	78

Abstract

Principles and Advances in Analysis of Ribonucleic Acid Sequence Using Nanopores

By

Andrew M. Smith

This work describes advances in nanopore sequencing technology as they apply to RNA. RNAs in the cellular environment have a wide range of functions and structures, by nature being much more dynamic in their activities than their chemical counterpart, DNA. The diversity of activities and functions that RNAs assume in the cell is reflected in the diverse research interests of those who study RNA. Any effort to develop nanopore-based direct RNA sequencing applications requires accounting for this diversity.

Given the spectrum of interests in RNA biology, much of this dissertation attempts to address a fundamental challenge in directly analyzing RNAs by nanopore - how to deliver and read diverse classes of RNA at single-nucleotide resolution using a nanopore sequencer? The introductory first chapter describes background on nanopores as sequencing sensors and samples of some biology surrounding this diverse class of molecules. The state of the art and challenges in RNA sequencing in general are discussed.

The second chapter of this work demonstrates a method specifically preparing tRNA for nanopore sequencing. The primary challenge in this effort is to be able to consistently load a single tRNA end and process the tRNA strand through the pore in a linear order. A molecular adapter composed of double-stranded DNA ligated to the tRNA termini facilitates this process. The adapter specifically targets tRNA by hybridizing to the universally conserved CCA tail of tRNA. It also provides a binding

site for $\phi 29$ DNA polymerase, which acts a molecular brake on RNA under non-catalytic conditions. These two features allow for discrimination between two tRNA species from *E. coli*. This work demonstrates that it is possible to use a nanopore to analyze individual tRNAs as linear strands, which is a necessary prerequisite for nanopore-based tRNA sequencing.

The third chapter of this work describes applying the Oxford Nanopore MinION sequencer to directly sequence 16S ribosomal RNA (16S rRNA). Demonstration of direct RNA sequencing of poly-adenylated RNA by Oxford Nanopore in August of 2016 was a critical development towards direct sequencing of RNA in general. The work described in Chapter 3 again involves designing an adapter, but this time specific for prokaryotic 16S rRNA. This adapter can be used with the existing Oxford Nanopore direct RNA sequencing kit. Sequencing of 16S rRNA from *E. coli* and other three other microbial species is demonstrated. The data and confirmatory experiments show that it is possible to directly detect modified ribonucleotides in nanopore-based sequencing data, which has long been a promised benefit of direct RNA sequencing methods. The 16S rRNA adapter is designed to hybridize to conserved 3'-end sequence of *E. coli* 16S rRNA. It could be generalized to all prokaryotic 16S rRNA, which would be desirable for rapid identification of prokaryotic microbes in a clinical or environmental setting.

The appendices cover unpublished work on helicase proteins to control RNA strand movement through a nanopore. The appendices also contain protocols for preparing different classes of RNA for nanopore sequencing and preparation of the helicase fragment of the SV40 Large T-antigen.

Dedication

To my spouse and partner, Morgan, and my family. I love you all.

Acknowledgements

This work would not have been possible without many individuals providing both intellectual support and encouragement. It is worth mentioning a few of these individuals, but this acknowledgement is in no way comprehensive or does justice to the investment these individuals made in my personal growth. Mark Akeson, as my dissertation advisor, guided my scientific inquiries and trained me to design well-conceived scientific experiments. His guiding principles for scientific communication have sharpened my writing and will continue to drive me towards always improving. David Bernick has shown me how it is possible to traverse across scientific fields, all the while pursuing questions that most deeply engage one's mind. He has been an invaluable resource on many topics and has also helped me keep my sense of humor through this process. Probably most importantly, he believed that I was capable of a body of research at the caliber necessary to complete a Ph.D, which kept me pursuing that goal. There have been many helpful conversations with fellow lab members and neighbors: Hugh Olsen, Kate Lieberman, Robin Abu-Shumays, Miten Jain, Art Rand, Logan Mulroney, Max Genetti, Stas Fridland, Arian Mackie, Colin Naughton and Stas Federicin. I would particularly like to thank Robin for always being supportive, having a positive outlook and for uncounted hours of assistance in collecting data from single-channel nanopore experiments. I thank Art for always challenging me. His drive to pursue the highest quality science has pushed me to be a better scientist. I thank Logan for many helpful science conversations and always being an ally. His friendly and open attitude has allowed me to develop my scientific voice. I thank Miten for endless hours of conversation and brainstorming. His broad knowledge has enriched my thinking and his creativity has made me a more imaginative scientist. I am also thankful to other research groups who have provided knowledge and support: the Noller Lab, the Lowe Lab, and the Ares Lab.

Chapter 1

Introduction

Background on nanopore sequencing

Nanometer-scale pores can be found in nature, formed by biological protein subunits. Monomers of these porin proteins, such as α -hemolysin from the bacterium *Staphylococcus aureus*, assemble to create highly reproducible pore structures in biological membranes (1). The idea that these nanometer-scale structures could be used as biosensors arose in the early 1990's, with the conception occurring in two distinct academic groups from Harvard and UC Santa Cruz (2). The concept was simple – an ionic current in a buffered solution could be measured through these pores if they were embedded in a membrane while a voltage potential was applied. Ionic polymers, such as DNA or RNA, would be drawn through the pore by the applied voltage. As the polymer passed through the pore, each nucleotide would influence the ionic current differently depending on the identity of the base, thus allowing sequence information to be decoded from the ionic current pattern. The fundamental experiment in 1996 by Deamer, Branton and co-workers demonstrated the single-file translocation of single-stranded DNA through an α -hemolysin nanopore was possible (3). A number of significant advances have had to occur since these initial results, but the principle of using ionic current to detect single stranded nucleic acids has ultimately evolved into a functional nanopore DNA sequencer, with the launch of Oxford Nanopore Technologies' MinION sequencer in 2014. While work continues on solid-state pores (4), these will not be discussed since there are no published results to-date demonstrating their utility for analyzing RNA or DNA sequence at single-nucleotide resolution.

Sequencing of RNA molecules directly with a nanopore has lagged behind the development of nanopore DNA sequencing, despite early efforts. The earliest published work from 1999 showed that RNA, like DNA could be translocated through an α -hemolysin nanopore, and that block homopolymers of adenosines and cytidines could be distinguished by their unique ionic current signatures (5). However, this work also revealed that uncontrolled nucleic acid strand movement proceeded at a rate that is too rapid (1-22 μ s/nt) to resolve individual nucleotide sequence from ionic current. The observed different ionic signatures for homopolymer As and Cs primarily arose as a result of stacked structures formed by these homopolymers in solution. Achieving control over strand movement through a pore poised a second fundamental challenge to using nanopores as sequencing tools.

A number of ideas to solve this problem have been proposed including: changing solution viscosity, using complementary oligonucleotides that must be stripped off during strand translocation or even coupling exonucleases to sequentially digest DNA proximal to the pore, allowing interrogation of mononucleotides (6). Another approach, using catalytic DNA processing proteins to control strand movement, such as polymerases or helicase, was initially described in 2007 (7). The first successful demonstration of the idea, published in 2010 (8), showed that a DNA polymerase could in fact act as molecular motor atop nanopores, with their atomic-scale catalytic cycle providing the necessary means to control a DNA strand in a nanopore. Between 2010 and 2012, the ϕ 29 DNA polymerase was used to demonstrate robust control of DNA translocation through a nanopore (9), with multiple rounds of catalysis resulting in successive single-nucleotide advancements along the DNA strand being interrogated in the pore (10, 11). In these experiments, DNA was being pulled in the opposite direction of the applied voltage by the enzyme. This resulted in the enzyme advancing along the strand under load, with registry errors along the template being relatively high. The DNA helicase Hel308 has since

been shown to perform a similar function, advancing along a strand being held under load in a nanopore (12). While no published data exists, there is evidence (this work, Appendix A) that a helicase can act in concert with applied voltage and still control nucleic acid strand movement in single nucleotide increments. It is generally thought that Oxford Nanopore Technologies sequencers operate with helicases as molecular motors in this fashion.

Oxford Nanopore Technology publicly announced the MinION sequencer for DNA in 2012 and provided a commercial device by late 2014. The MinION is a small (~90g), USB-enabled nanopore sequencer with an array of 2048 single pores controlled by 512 individual voltage-clamped amplifiers (13). This allows hundreds of single-molecule sequencing reactions to proceed at hundreds of bases per second. To prepare target DNA for sequencing, synthetic DNA strands pre-loaded with ONT's DNA motor protein are covalently ligated to target DNA ends. These adapters initiate threading of ssDNA through the pore and concentrate target DNA at the membrane via a hydrophobic group covalently attached to the adapter DNA. When coupled with a personal computer, the MinION is currently capable of generating $>10^9$ bases of sequence data over a 48 hour run (14) with sequence read identities averaging 92% (15). Read lengths are typically 8-10 kilobases (kb), with individual reads of hundreds of kilobases possible (15). It is likely the read lengths are primarily limited by DNA handling during extraction/library preparation and that reads will exceed 1 megabase (mb) in short order. The rapid advancement of nanopore DNA sequencing length and accuracy has been facilitated by the previously mentioned advancements in DNA processing enzymes, pore proteins (16, 17), and biomimetic membrane engineering (18).

It was presumed that nanopore direct RNA sequencing would rapidly follow on the advances in nanopore DNA sequencing, starting with demonstration of a suitable RNA processing protein for controlling RNA movement through a nanopore. This turned out not to be the case, and no suitable RNA motor protein has been

publicly described in detail. However, two alternative approaches were described in literature prior to the public release of ONT's direct RNA method in 2017. In 2012, Ayub and Bayley showed that a modified α -hemolysin nanopore could discriminate among ribonucleotides when they were held statically in the pore (19). As with DNA, they conceived that a RNA exonuclease coupled with a nanopore could sequence RNA if a number of challenges, including engineering the exonuclease-coupled nanopore and ensuring diffusion of rNMPs through the pore, could be met (20). The other approach, published in 2015, attempted to use the robust binding of ϕ 29 DNA polymerase to DNA to slow translocation of DNA-RNA chimeric molecules through α -hemolysin nanopore. The author and colleagues undertook this work in 2013-2014. That concept is described in detail in Chapter 2 of this dissertation. Ultimately, these two concepts are superseded by a true nanopore direct RNA sequencing method from Oxford Nanopore Technologies, which was released publicly for use in April of 2017. Their direct RNA sequencing kit incorporates all of the necessary molecular biology components to provide single nucleotide resolution on RNA strands.

Background on RNA sequencing

RNA sequence analysis has long been a component of RNA biology research in general. Pre-dating the development of high-throughput sequencing, microarrays allowed for large-scale interrogation of RNA transcripts. However, they are limited in their sequence resolution and dynamic range (21). Conventional high-throughput cDNA sequence sequencing, dubbed RNA-Seq (22), uses sequencing-by-synthesis (SBS) chemistries to read cDNA. It is a well-established tool and has advanced our understanding of RNA biology on a genome-wide scale (23). Short read sequencing is the most widely used and generally offers reads from 50-250 bases in length. Library preparations for these methods require target-specific or random primed

reverse transcription to generate cDNA, followed by fragmentation and PCR amplification. Because double-stranded cDNA is created and sequenced in these protocols, special efforts must be made to retain information on which strand the original RNA transcript is derived from. More recently, long transcript reads (>1kb) are obtained using the Iso-Seq method from Pacific Biosciences (24). However, this method also sequences cDNA after reverse transcription, which ultimately destroys any post-transcriptional marks on the original RNA.

To address this issue, a number of existing chemical or molecular biology methods for detection of RNA modifications have recently been coupled with high-throughput RNA-seq. These methods allow inference of RNA modification sites on a transcript-wide scale. Methods are described that can detect 6-methyladenine (m6A), 1-methyladenine (m1A), pseudouridine (Ψ), 5-methylcytosine (m5C), and inosine (I), among others (25) (**Figure 1.1a**). Discoveries resulting from these methods have provided new insights into the dynamics of RNA modifications and resulted in renewed interest in the role that RNA modifications play in biology.

The methodologies for these techniques vary. Given the large body of literature on the use of chemical reagents with RNA (26), a number of approaches are based on selective chemical reactions with modified nucleotides. Historically, these types of reactions would have been coupled with primer extension to analyze a single RNA species, but now they are being applied on a transcriptome-wide scale. Pseudouridine is detected using a well-established chemical approach. The chemical reagent 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate (CMCT) selectively reacts with Ψ residues and is followed by secondary treatment to remove CMC adducts from off-target reactions with U and G residues. The RNA is then reverse transcribed, which stalls at CMC adducts, and cDNA is sequenced (27).

Coupled with RNA-seq, transcript-wide pseudouridine content is inferred from reverse transcription stops not observed in an untreated control sample. Detecting m6A uses antibody-based immunoprecipitation (IP) enrichment of m6A-containing RNA fragments followed by reverse transcription and cDNA sequencing (28). Motif sequence analysis is used to predict m6A sites, identifying methylation consensus motifs in reads enriched relative to a control IP sample. A version of this method that allows single-nucleotide resolution of m6A was recently described (29). In either case, reverse transcription is used and the presence of RNA modifications is inferred from cDNA sequence data.

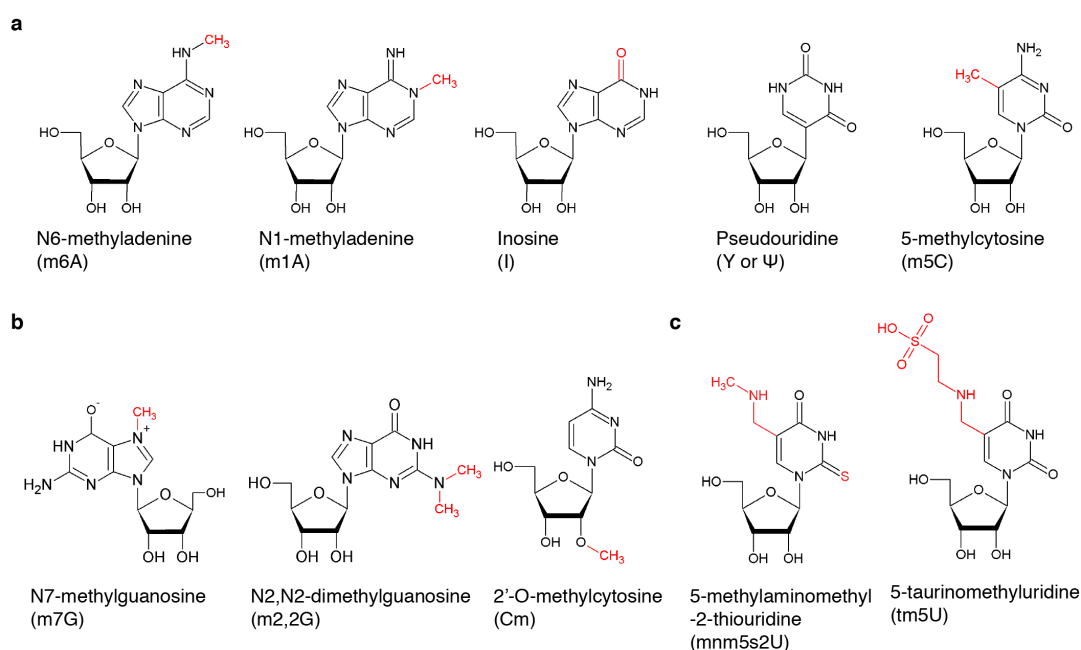


Figure 1.1: Some modifications commonly found in RNA

(a) Chemical structures of nucleoside modifications commonly found in mRNA. The chemical groups added by the modification are indicated in red, except for pseudouridine. Pseudouridine is an isomer of uridine, but enzymatic rotation of the uridine base results in an additional hydrogen bond donor at position 5 on the pyrimidine ring. (b) Some modifications found in rRNA and other classes of RNA. (c) Uracil modifications found in the first anticodon position of tRNA. These modifications require multiple enzymes to complete. They are necessary for efficient wobble base-pairing between the tRNA and mRNA codon. The structures are from the Modomics website, <http://modomics.genesilico.pl>

These methods are predominately focused on modifications found in messenger RNA and poly-adenylated transcripts. There are over 100 known RNA modifications (30), and other classes of RNA, such as rRNA and tRNA, contain a more diverse set of modifications. In addition to the previously mentioned modifications, ribosomal RNAs contain modifications such as N²,N²-dimethylguanosine (m²,2G) and 2'-O-methylation (**Figure 1.1b**). Individually, all of the modifications in *E. coli* ribosomal RNA are dispensable, and are thought to be primarily involved in optimizing ribosomal subunit tertiary fold or catalytic activity (31). tRNAs contain probably the greatest diversity of modifications, typically containing 5-7 modifications per tRNA species. Some of these are simple and stabilize the tRNA tertiary fold, such as m¹A has been shown to do (32). Other modifications are more complex, added post-transcriptionally by a multi-enzyme process (33). 5-taurinomethyluracil (tm⁵U) and 5-methylaminomethyl-2-thiouridine (mnm⁵s²U) are two important examples (**Figure 1.1c**). These modifications start with uracil found at wobble position of the tRNA anticodon (position 34 by canonical tRNA numbering). This position is modified by at least two enzymes, hGTPBP3 and hMTO1 in the case of tm⁵U in humans (34, 35). These two complex uracil modifications stabilize tRNA wobble pairing with A or G in the third position of mRNA codons (36, 37). Other modifications, such as 5-formylcytosine (f⁵C) and 5-taurinomethyl-2-thiouracil (tm⁵s²U), are also found at the first position of other tRNA anti-codons and serve similar roles in stabilizing wobble codon recognition (38, 39). The importance of nucleotide modifications in tRNA anticodons is

underscored by the finding that loss of tm5U modification in human mitochondrial tRNA(Leu) is pathogenic. It results in a defective tRNA recognition of UUG codons (40). The resulting translation defect in UUG-rich mitochondrial transcripts causes the rare heritable disease MELAS (mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes syndrome) (41). Similar rare conditions are associated with loss of other tRNA wobble modifications (42–44). Given the diversity and complexity of RNA modifications, it seems both unlikely and impractical that all modifications will have high-throughput RNA-seq methods developed to detect them.

Besides requiring indirect methods to identify post-transcriptional modifications, SBS sequencing of cDNA has limitations such as: short reads ambiguously mapping to multiple loci, difficulty in resolving alternative splicing and quantifying isoform abundance, reverse transcription errors, and PCR amplification bias. Because of these caveats, the concept of direct, high-throughput RNA sequencing has long held appeal. Two different attempts are notably found in literature, both by commercial companies. The first was from Helicos in 2009, which used fluorescence-based nucleotides incorporated during cDNA synthesis to infer RNA sequence without an intervening library preparation (45). The second approach came from Pacific Biosciences, and used a similar chemistry and a unique imaging method involving zero mode wave-guides (46). Both these methods were limited by short reads and low throughput. Neither of these results led to a widely accepted direct RNA sequencing methodology that addressed the issues inherent to SBS.

Oxford Nanopore Technologies described a nanopore-based direct sequencing RNA method in August 2016 (47). It uses very different principles to sequence RNA than the previously described methods and offers benefits that conventional high-throughput SBS cannot provide. With nanopore direct RNA sequencing, each original RNA strand is read as a linear molecule passing through the nanopore. This means

reverse transcription and PCR can be omitted and no additional RNA fragmentation is necessary. This makes it possible to read full-length transcripts, which should aid in detecting and quantifying RNA splice isoforms.

Because nanopore RNA sequencing directly reads the RNA strand, simultaneous detection of different modifications is also a possibility and would likely require no additional preparation. This would permit examination of interplay between different post-transcriptional modifications. Discovery of unknown modifications is also possible when reading RNA *ex vivo*. This will require identifying anomalous ionic current patterns in the raw data and orthogonal methods, such as mass spectrometry, to identify the modification. Ultimately, RNA modifications will be able to be identified and quantitated on a per strand basis at single-nucleotide resolution from nanopore data, similar to work already being undertaken with DNA modifications m5C and m6A (48).

Outline of the Dissertation

The chapters that follow are comprised of two research projects that are the subject of manuscripts either published or under review. Chapter two describes a project aimed at preparing tRNA for direct RNA sequencing. This work was undertaken prior to the development of the ONT direct RNA sequencing method. All data was collected from single-pore/bilayer experiments. The set of experiments combined to show that it is possible to covalently add adapters to tRNA and draw these highly structured RNAs through a nanopore. Further experimentation shows that even the relatively insensitive α -hemolysin nanopore is sufficient to differentiate between two tRNA with different sequences. A manuscript based on this work was published in *Frontiers in Bioengineering and Biotechnology* in June of 2015.

Chapter three describes customization of the ONT direct RNA sequencing kit for 16S ribosomal RNA. As a phylogenetic marker and a target of antibiotics, the 16S rRNA is a logical choice to target for selective, direct RNA sequencing. Direct

sequencing of *E. coli* 16S rRNA is described, and the customized adapter can be modified to accommodate small variations in 16S rRNA 3' ends. Using a genetic approach, the work also demonstrates that RNA modifications m7G and pseudouridine can be discriminated from their canonical counterparts. It seems likely that direct sequencing of 16S rRNA will be useful as a rapid, point-of-care diagnostic tool. A manuscript from this work was submitted for review to *Nature Biotechnology* in April of 2017.

The appendices describe unpublished work on helicases to control RNA strand movement through a nanopore, continuing work on achieving single-nucleotide resolution on tRNA, and protocols used to prepare RNA for nanopore analysis.

Individual Contributions

The project in Chapter 2 was conceived of by the author, Mark Akeson and David Bernick. The experiments were designed by the author, Mark Akeson and David Bernick. RNA bench experiments were performed by the author. Nanopore experimental data was collected and analyzed by the author and Robin Abu-Shumays. The manuscript was co-written by the author, Mark Akeson, Robin Abu-Shumays and David Bernick.

The project in Chapter 3 was conceived the author. RNA bench experiments were conceived and performed by the author and Logan Mulroney. MinION experiments were designed by the author, Miten Jain, and Mark Akeson. Data analysis was performed by the author, Miten Jain and Daniel Garalde. The manuscript was co-written by the author, Miten Jain, Daniel Garalde, and Mark Akeson.

The project in Chapter 4 was conceived by the author. The author and Max Genetti performed bench experiments. Single-channel nanopore data was collected by

the author and Robin Abu-Shumays. Data was analyzed by the author, Robin Abu-Shumays, and Max Genetti. The text and figures were written by the author. The results of Chapter 4 have not been published or presented in any conference.

Chapter 2

Capture, unfolding and detection of tRNA using a nanopore device

Introduction

Transfer RNAs (tRNA) decode genetic information, delivering to the protein-synthesizing ribosome the individual amino acids specified by each codon of a messenger RNA. In light of this, it is unsurprising that they are the most numerous RNA species in the cell, composing approximately 80% of the RNA molecules per generation in yeast (49). These small, non-coding RNA molecules contain numerous post-transcriptionally modified nucleotides, which contribute to tRNA fold stabilization, codon recognition, and aminoacylation [reviewed in (50) and (32)]. A growing body of research indicates that tRNAs and their nucleotide modifications are directly targeted for regulation as well as acting as potential regulators themselves [reviewed in (51) and (52)].

Current methods for analysis of tRNAs, and RNA in general, include RNA-seq, microarray, and mass spectrometry. These methods are proven tools for detection of novel tRNAs and global tRNA expression patterns (53, 54). However, each method has limitations.

High-throughput RNA sequencing (RNA-Seq) methods require extensive library preparation, including PCR amplification, to prepare cellular RNA for sequencing. A reverse transcription (RT) step is necessary to copy the original RNA sequence to cDNA, which results in loss of the original RNA strand. Additionally, the RT step is impeded by the occurrence of structure and nucleotide modifications, which are both commonly found in tRNAs. These “RT-stops” result in truncated

cDNA. While these RT stops have been exploited by various methods to infer structure or nucleotide modification state in RNA, they do not permit direct detection of nucleotide sequence and modification identity along intact RNA (46, 55–57). Methods have been developed for detecting a few specific RNA modifications, such as bisulfite treatment for 5-methylcytosine or immunoprecipitation for N6-methyladenine, but these require additional labor-intensive steps (28, 58).

Microarrays share a similar problem with loss of modification information and extended processing times. Success has been reported in detecting nucleotide modifications in tRNAs that affect hybridization of the target molecule to array probes, which limits detection to those modifications that affect Watson–Crick base-pairing (59, 60). In addition, microarray probe design requires prior knowledge of the target sequence, which limits detection to known RNAs.

Liquid chromatography coupled to mass spectrometry (LC- MS) has also been applied to study tRNAs. In particular, two recent publications have used LC-MS to examine dynamic changes in tRNA modification state in *Saccharomyces cerevisiae* under varied environmental growth conditions (61, 62). However, the nucleolytic fragmentation that is required by this method prevents observing modifications in the full sequence context of the tRNA.

Detection of tRNA subpopulations that result from variability in modification, as seen in mitochondrial diseases, will benefit from direct interrogation of intact, single tRNA molecules (40). In two specific mitochondrial examples, failed tRNA modifications cause dysfunctional translation of mitochondrial enzymes. These in turn result in the mitochondrial encephalopathies, myoclonic epilepsy with ragged-red fibers (MERRF), and mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) (41, 44). In these examples, a method that examines individual tRNA molecules could reveal the extent of incompletely modified or mutant forms in these complex disease phenotypes.

Nanopore sensors interrogate single molecules and should allow for examination of several thousand individual tRNA in a single experiment. Nearly 20 years ago, biological nanopores with 1–2 nm limiting apertures were conceived as single-molecule sensors for nucleic acids (3). In concept, the nucleotide sequence of an individual molecule would be read by observing changes in ionic current as the linearized strand is electrophoresed through the nanopore aperture. Recent developments in sensing DNA have coupled an enzyme to regulate DNA movement in single nucleotide steps through a nanopore, which produced ionic current traces that provide a single base readout of DNA sequence (10, 11). Additionally, experiments by our group and others have shown that DNA cytosine modifications can be detected with high confidence from individual nanopore reads of chemically synthesized DNA (63, 64). Applying these principles should provide a high-throughput method to directly examine tRNA nucleotide sequence.

With RNA, strand sequencing with single-nucleotide resolution has yet to be demonstrated with a nanopore device. However, Ayub and Bayley have shown that immobilized RNA strands within a modified α -HL nanopore apparatus yield distinct ionic current amplitudes that discriminate between canonical and select modified ribonucleotides (19). Further, solid-state nanopores with pore-sizes that exceed 3nm have been shown to distinguish folded tRNA molecules from linear double-stranded RNA and DNA (4). By extension, these results suggest that nanopore sensors could detect sub-molecular features of tRNA, including nucleotide modifications, if tRNAs can be mechanically unfolded and electrically motivated to pass through the pore.

With this in mind, we sought to develop a method to specifically capture tRNA molecules, promote their mechanical unfolding, and initiate threading of the linearized strand through the nanopore lumen. To this end, we designed an oligonucleotide adapter that can be attached to intact tRNAs. To slow tRNA translocation through the pore, we employed a non-catalytic protein “brake” that loads onto the adapter. This allowed us to determine the direction of strand

translocation and provided sufficient resolution to determine ionic current signal features associated with the translocating adapter and the tRNA. Results presented here demonstrate that tRNA attached to such an adapter and modulated by a protein brake can be completely translocated through the α -HL nanopore, and that *Escherichia coli* tRNA^{fMet} and tRNA^{Lys} produce differentiable nanopore signals in this system. This provides the foundation for future work aimed at achieving single nucleotide resolution of individual transfer RNA.

Results

Capture and threading of tRNA through the α -HL nanopore is facilitated by ligation of an oligonucleotide adapter to the tRNA

Reading the nucleotide composition of individual tRNA molecules will require capture, denaturation and threading of each strand sequentially through the nanopore. In our initial experiments, we found that native tRNA molecules caused long (>30 seconds) ionic current blockades of the α -HL pore (data not shown). These molecules had to be ejected by voltage reversal to re-establish an open-pore ionic current. This suggested that tRNA molecules in their native form would not readily translocate through the α -HL nanopore.

We reasoned that an extended single-stranded region, longer than the ACCA in native tRNA, may be needed to initiate threading of each tRNA molecule into the lumen of α -HL. To accomplish this we devised a strategy to covalently attach synthetic nucleic acid strands to the 3' and 5' ends of the tRNA. This was achieved using a *Y-shaped*, partially double-stranded DNA-RNA adapter that contained a 3' RNA overhang complementary to the universally conserved CCA tail in tRNA

(**Figure 2.1**). The strand of the adapter, which bares the 3' RNA overhang, was designed to be ligated to the 5' end of a tRNA (referred to as the “leading strand”). The unpaired region of the leading strand contained 35 single-stranded nucleotides and was designed to facilitate capture and threading into the nanopore. The double-stranded region of the adapter was designed to allow a dsRNA ligation, and effectively extended the tRNA adapter stem by 15 base-pairs.

The strand of the adapter that was designed to be covalently attached to the 3' end of the tRNA (referred to as the “trailing strand”) incorporated a cholesterol tag at its 3' end. This was designed to locally concentrate the adapted tRNA at the lipid bilayer-aqueous interface of the nanopore experimental setup. Association of the cholesterol moiety on the 3' trailing strand with the bilayer should favor capture of the free 5' end of the leading strand in the electric field surrounding the nanopore. Finally, the adapter design incorporated abasic residues into both leading and trailing strands to act as ionic current signal markers upstream and downstream of the ligated tRNA.

We initially tested if the adapter could be enzymatically ligated to tRNA using T4 RNA Ligase 2 (RNL2). Analysis of that ligation reaction revealed that a product of appropriate size (~160 nt) was generated only in the presence of both the adapter and a model tRNA substrate (*S. cerevisiae* tRNA^{Phe}) (**Figure 2.1b**). These results indicated that enzymatic ligation with RNL2 was an effective method for adding adapter strands to this model tRNA. To test this further, we ran additional ligation experiments with *E. coli* tRNA^{fMet}. This tRNA represented a more challenging substrate because it contains a non-canonical nucleotide pair at the end of the acceptor stem (65). Our results showed that *E. coli* tRNA^{fMet} was also a reactive substrate for ligation to the adapter (data not shown).

Initial nanopore experiments were performed with adapted *E. coli* tRNA^{fMet} using an established single-channel apparatus (see Methods) (**Figure 2.2A**). Ionic current blockade events were observed with a typical duration of tens of milliseconds (mean duration = $10^{-2.6 \pm 0.06}$ seconds, variation of the mean shown as SEM)(**Figure 2.2b** and **2.2c**, magenta circles). These events were longer than the events observed for the adapter alone, which were typically a millisecond or less (mean duration = $10^{-3.9 \pm 0.04}$ seconds) (**Figure 2.2c**, open triangles). The increased duration of events with adapted tRNA^{fMet} suggested that longer or more structured molecules were being captured and translocated through the pore. The extremely short duration blockade events observed with the adapter alone were consistent with single-stranded nucleic acids being translocated through the pore (66).

If the longer dwell times were caused by tRNA entering the nanopore, then conditions known to stabilize tRNA should increase event duration further. Magnesium ions are known to stabilize the tertiary fold of tRNA (67, 68). When we added magnesium chloride to the experimental buffer (5mM final concentration) we observed an increase in event duration (100 fold or 2 Log₁₀ units) relative to experiments absent Mg²⁺ (mean duration = $10^{-1.4 \pm 0.1}$ seconds) (**Figure 2.2c** blue squares). These blockades were self-terminating, as seen in experiments without Mg²⁺. This result was consistent with capture of a Mg²⁺-stabilized tRNA. Together, these results suggested that the electric-field driven denaturation of secondary, and potentially tertiary structure, facilitated the translocation of tRNA through the pore.

φ29 DNAP acts as a molecular brake during translocation of adapted RNA under non-catalytic conditions

In previous studies with both RNA and DNA, uncontrolled polynucleotide translocation rate through a nanopore was too high to resolve single nucleotide-level information about the translocating strand (66). Therefore we sought to slow translocation of adapted tRNA to improve resolution of tRNA features and to provide

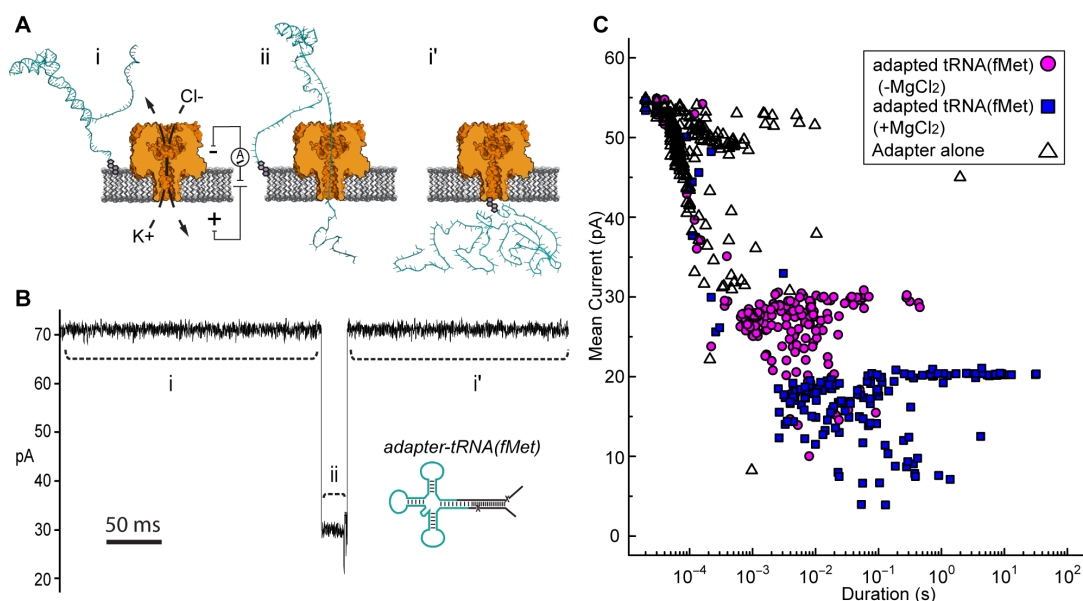


Figure 2.2: Adapted tRNA-dependent ionic current blockades observed during single-channel α HL nanopore experiments.

(A) cartoon illustration of the single-channel nanopore apparatus and a proposed adapted tRNA translocation event. i) A constant voltage (trans side +) is applied across a single α -HL nanopore (orange) embedded in a lipid bilayer (grey). ii) Electrophoretic capture of an adapted tRNA (cyan) results in a decrease in the measured ionic current through the nanopore. i') Return to open channel current when the tRNA clears the pore in the trans compartment. (B) An ionic current trace from a nanopore experiment with adapted *E. coli* tRNA^{fMet} (inset). Ionic current regions i-ii and i' in the trace (dashed lines) correspond to the proposed tRNA translocation event in panel A. The blockade event shown is typical of thousands of events observed during nanopore experiments with adapted tRNA^{fMet}. Scale bar indicates 50 milliseconds. (C) Nanopore blockade mean ionic current versus duration caused by adapted tRNA^{fMet} in the presence or absence of Mg^{2+} . The mean current and duration of approximately two hundred events are shown for representative nanopore experiments with either adapted tRNA^{fMet} (- Mg^{2+}) (magenta circles) or adapted tRNA^{fMet} (+ Mg^{2+}) (blue squares). The adapter on its own (- Mg^{2+}) (open triangles) was also examined as a negative control. In all cases, single-channel α -HL nanopore experiments were conducted at 180mV (trans side +) with tRNA substrate at 0.5nM in 0.3M KCl, 10mM HEPES (pH 8.0), and +/- 5mM MgCl₂ (see Methods).

definitive evidence that adapted tRNA molecules transit the pore in their entirety. Control of DNA transit rates using DNA polymerases has been documented, and Lieberman et al. (2010) showed that $\phi 29$ DNAP can serve as a ‘molecular brake’ that controls the rate of DNA translocation through the α -HL pore under non-catalytic conditions (absent Mg^{2+} and dNTPs) (9). Furthermore, this molecular brake activity of $\phi 29$ DNAP has been observed on a chimeric DNA-RNA substrate with a nanopore device (J. Clarke, Oxford Nanopore Technologies, pers. comm).

We wanted to determine if the $\phi 29$ DNAP molecular brake could also be used to control translocation of RNA molecules containing more complex structures, such as stem-loops found in tRNA. For this we synthesized a simple RNA hairpin, which we ligated to the tRNA adapter (**Figure 2.3a**). This synthetic RNA (referred to as RNA(hairpin)) mimicked the acceptor stem of tRNA^{Met}, where the two halves of the acceptor stem were linked by a short loop region of five uracil residues.

Nanopore capture of complexes formed between the adapted RNA and $\phi 29$ DNAP, similar to those seen by Lieberman et al., should result in greatly increased dwell time of individual adapted RNA molecules within the aperture of α -HL. This would be observed as population of longer duration nanopore current blockades, which would be distinct from the shorter duration events of unbound adapted RNA strands.

Control nanopore experiments with the adapted RNA(hairpin) construct absent $\phi 29$ DNAP resulted in current blockade events with mean duration on the order of milliseconds (mean duration = $10^{-3.2 \pm 0.02}$ seconds) (**Figure 3c**, magenta circles). Addition of $\phi 29$ DNAP to the buffer solution containing adapted RNA(hairpin) on the *cis* side of the nanopore apparatus produced two different types of current blockade events. These events typically fell into one of two populations:

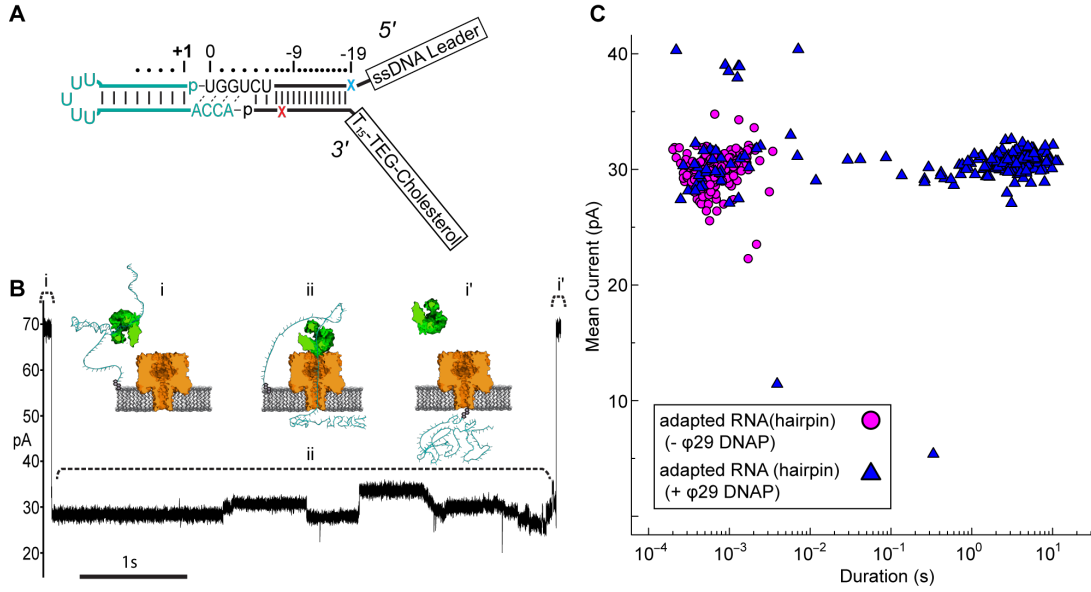


Figure 2.3: Nanopore capture of adapted RNA complexed with non-catalytic ϕ 29 DNAP

(A) cartoon illustration of the single-channel nanopore apparatus and a proposed adapted tRNA translocation event. *i*) A constant voltage (*trans* side +) is applied across a single α -HL nanopore (orange) embedded in a lipid bilayer (grey). *ii*) Electrophoretic capture of an adapted tRNA (cyan) results in a decrease in the measured ionic current through the nanopore. *i'*) Return to open channel current when the tRNA clears the pore in the *trans* compartment. (B) An ionic current trace from a nanopore experiment with adapted *E. coli* tRNA^{fMet} (inset). Ionic current regions *i-ii* and *i'* in the trace (dashed lines) correspond to the proposed tRNA translocation event in panel A. The blockade event shown is typical of thousands of events observed during nanopore experiments with adapted tRNA^{fMet}. Scale bar indicates 50 milliseconds. (C) Nanopore blockade mean ionic current versus duration caused by adapted tRNA^{fMet} in the presence or absence of Mg²⁺. The mean current and duration of approximately two hundred events are shown for representative nanopore experiments with either adapted tRNA^{fMet} (-Mg²⁺) (magenta circles) or adapted tRNA^{fMet} (+Mg²⁺) (blue squares). The adapter on its own (-Mg²⁺) (open triangles) was also examined as a negative control. In all cases, single-channel α -HL nanopore experiments were conducted at 180mV (*trans* side +) with tRNA substrate at 0.5nM in 0.3M KCl, 10mM HEPES (pH 8.0), and +/- 5mM MgCl₂ (see Methods).

a short duration population (duration <0.1 seconds, mean $10^{-2.9 \pm 0.09}$ seconds), similar to events in the control experiment, and a long duration population not seen in the control experiment (duration ≥ 0.1 seconds, mean $10^{0.43 \pm 0.03}$ seconds) (**Figure 2.3b** and **2.3c**, blue triangles). The shorter duration population (<0.1 s) appeared consistent with RNA(hairpin) absent $\phi 29$ DNAP and was statistically indistinguishable from the event population seen in the control (p -value < 0.66 , 2-tailed T-test). The longer duration population (≥ 0.1 s) was longer in mean duration and was statistically different from the population seen in the control (p -value < 0.0001 , 2-tailed T-test). This suggested that these long duration events were the result of $\phi 29$ DNAP binding the adapted RNA substrate and slowing strand translocation through the nanopore.

We included two abasic residues (1'-H deoxyribose) in the adapter strands near the ligation junctions with the RNA to act as indicators of strand translocation (see **Figure. 2.3a**). Abasic residues have been shown to cause distinctive high current states that are apparent during enzyme-controlled translocation of oligonucleotides through the α -HL pore (9, 69). Because the abasic residues in the adapter (subsequently referred to as a “dual-abasic adapter”) flank the RNA insert, they should translocate through the nanopore before and after the RNA insert (**Figure 2.4a**). This should produce an ionic current trace with high current states bracketing an intervening region and indicate strand translocation occurred in a linear conformation. Further, the intervening region would correspond to the adapted RNA(hairpin) insert traversing the pore. As seen with DNA during $\phi 29$ DNAP-mediated translocation, the individual ribonucleobases of the extended chain should directly modulate the ionic current during translocation through the α -HL pore, resulting in a reproducible pattern of ionic current states (10).

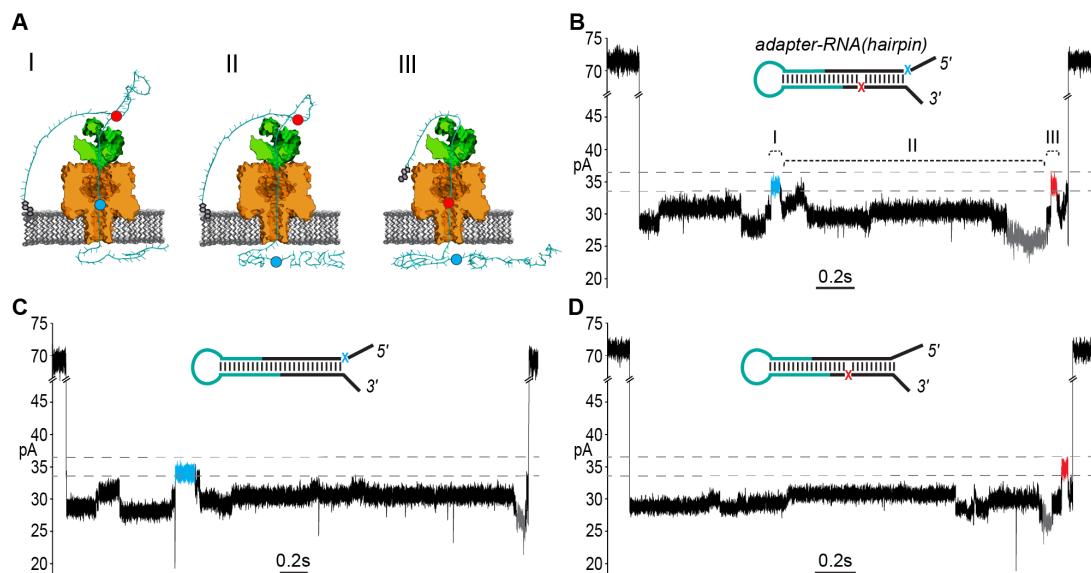


Figure 2.4: End-to-end translocation of adapted RNA(hairpin) constructs through α -HL nanopore.

(A) A cartoon model showing capture and translocation of an adapted RNA(hairpin) complexed with ϕ 29 DNAP. The adapted RNA(hairpin) substrate is described in Figure 3. *I*) Following capture of the 5' end of the adapter strand, the polynucleotide translocates through α -HL until the leading 5' abasic residue (blue circle) reaches pore limiting aperture. *II*) Translocation continues through the RNA portion of the adapted molecule. *III*) The trailing 3' abasic residue (red circle) reaches the pore limiting aperture after translocation of the RNA(hairpin). (B) A nanopore ionic current trace during translocation of an adapted RNA(hairpin) (inset), which corresponding to the cartoon in panel A. The event displays two high current marker regions characteristic of the 5' abasic residue (blue segment) and 3' abasic residue (red segment) built into the adapter. Translocation events that contained both leading and trailing markers were observed in ~27% of events with durations exceeding one second. The marker ionic current level was 33.5-36.5pA (dashed gray lines) under the experimental conditions used (see Methods). Scale bar indicates 0.2 seconds. (C) A nanopore ionic current trace during translocation of an adapted RNA(hairpin) (inset) bearing only the 5' abasic residue (5' mono-abasic adapted RNA). (D) A nanopore ionic current trace during translocation of an adapted RNA(hairpin) (inset) bearing only the 3' abasic residue (3' mono-abasic adapted RNA).

As predicted, translocation of dual-abasic adapted RNA(hairpin) constructs bound to ϕ 29 DNAP resulted in blockade events containing two distinct ionic current states in the range of 33.5pA to 36.5pA that flank a reproducible pattern of ionic current (**Figure 2.4B** and **Figure 2.7**). This suggested that the adapted RNA(hairpin) translocated completely through the nanopore, and that ϕ 29 DNAP acted as a passive molecular brake for both the DNA and RNA portions of the chimeric molecule.

The ionic current pattern produced by 5' and 3' abasic residues establishes the direction of translocation and indicates complete strand translocation

To establish the direction that ϕ 29 DNAP-bound strands translocated through the nanopore, it was necessary to assign each of the observed high current states to either the 5' or 3' abasic residue. To do this we synthesized adapters that contained only one of the 5' or 3' abasic residues. These “mono-abasic adapters” were ligated to the RNA(hairpin) substrate (see **Figure 2.4A**, cyan).

Substrates bearing the 5' mono-abasic adapter typically produced events containing a single high current state (33.5-36.5pA) near the beginning of the event (**Figure 2.4C**). Substrates bearing the 3' mono-abasic adapter caused a similar high current state near the end of the event, which was preceded by a low current state (\leq 26.5pA) (**Figure 2.4D**).

These observations provided an ionic current model for complete translocation of the dual abasic-adapted RNA(hairpin) in the 5'-to-3' direction. That is, the 5' abasic residue caused the first high current state, which was followed by the intervening current region corresponding to the RNA, which includes the low current

state. This is followed by the high current state proximal to the end of the event that is caused by the 3' abasic residue.

Using this model, we quantified the frequency of the leading and trailing high current states (henceforth referred to as leading and trailing markers) in events greater than one second duration. Translocation of the dual-abasic adapted RNA resulted in 27.4% of events (102 of 372) that contained both the leading marker and trailing marker separated by a region containing a low current segment (**Table 2.1**). An additional 43% of events were classified as containing only a leading marker and 6.7% were classified as containing only a trailing marker. The disparity in frequency between leading and trailing markers suggested that the 3' end of the strand was more difficult to resolve.

The less frequent observation of the trailing marker relative to the leading marker suggested two possible explanations: either the 3' abasic was absent from many strands, or that the translocation rate increased in the region of the 3' abasic. In experiments with dual-abasic molecules, where both 5' and 3' abasic markers were resolved, the duration of the 3' marker is notably shorter (**Figure 4B** and **Figure 7A**), supporting the latter hypothesis. Furthermore, in experiments with constructs containing mono-abasic adapters, we observed the same trends: strands bearing the 5' mono-abasic adapter produced events with a single high current region 64.5% of the time with notably longer duration (**Table 1**, **Figure 2.4C** and **Figure 2.7B**), and strands bearing the 3' mono-abasic adapter produced a smaller fraction of events (37.9%) with a single high-current marker of shorter duration (**Table 1**, **Figure 2.4D** and **Figure 2.7C**). This suggests that the translocation rate of the molecule increases after ϕ 29 DNAP unzips the double stranded region. This would produce shorter duration ionic current signatures at the 3' end. We reasoned that these shorter

duration 3' trailing markers were frequently unresolved at the measurement timescale used by our apparatus.

Table 2.1: RNA(hairpin) nanopore translocation events classified by detection of leading and trailing high current markers.

Substrate	Events	Leading and trailing high current observed ^a	Leading high current observed ^b	Trailing high current observed ^c	Other events ^d
RNA(hairpin) (dual-abasic adapter)	372	102 (27.4%)	160 (43.0%)	25 (6.7%)	85 (22.9%)
RNA(hairpin) (5' mono-abasic adapter)	287	3 (1.0%)	185 (64.5%)	10 (3.5%)	89 (31.0%)
RNA(hairpin) (3' mono-abasic adapter)	285	3 (1.1%)	4 (1.4%)	108 (37.9%)	170 (59.6%)

All events were ≥ 1 second duration

^a These events contained three required features as illustrated in Figure 4: 1) a high current marker segment (mean current 33.5-36.5pA with ≥ 2 milliseconds duration); 2) a low current segment (mean current ≤ 26.5 pA with ≥ 10 milliseconds); 3) a second high current segment proximal to the termination of the event.

^b These events contained: 1) a single high current marker segment (mean current 33.5-36.5pA with duration ≥ 2 milliseconds); 2) a low current segment (mean current ≤ 26.5 pA with ≥ 10 milliseconds) that always followed the high current marker.

^c These events contained: 1) a single high current marker segment (mean current 33.5-36.5pA with duration ≥ 2 milliseconds); 2) a low current segment (≤ 26.5 pA with a duration of 10 milliseconds) that always preceded the high current marker segment.

^d Events classified as "other" included all events that could not be assigned to one of the other categories, such as those displaying more than two high current marker states and events displaying no high current marker segments.

The result of the experiments using mono-abasic markers is consistent with 5'-to-3' traversal of adapted RNA through the α -HL pore when bound to ϕ 29 DNAP. By design, the 3' end of adapted RNA would be retained in or at the lipid bilayer. This would preferentially leave the 5' end of adapted RNA molecules available to initiate threading through the pore and act as the loading site for ϕ 29 DNAP. By experiment, when a single abasic residue was present only on the leading strand of the adapter (5'), we observed a single long-duration high current marker towards the beginning of ϕ 29 DNAP-bound translocation events. Conversely, inclusion of a single abasic residue only on the trailing strand of the adapter (3') produced a single short-duration high current marker towards the end of ϕ 29 DNAP-bound translocation events. Presence of both abasic residues in the adapter resulted in ~27% of ϕ 29 DNAP-bound translocation events containing two high current markers. Taken together, these results support initiation of the ϕ 29 DNAP-bound strand threading through the nanopore from the 5' end, followed by the RNA portion, and terminating with the 3' end of strand passing through and exiting the pore. A 5'-to-3' voltage-driven “unzipping” process was previously observed for ϕ 29 DNAP under non-enzymatic conditions on DNA (70). We conclude from the results presented here that: ϕ 29 DNAP is being driven in the 5' - 3' direction, it unzips the dual-abasic adapted RNA(hairpin) strand, and in events where both markers were observed, the strand is translocated through the nanopore in its entirety. Further, these markers provide approximate boundaries of the RNA-dependent portion of the nanopore signal.

***E. coli* tRNA^{fMet} and tRNA^{Lys} can be classified based on their nanopore current signals**

If the ionic current segment flanked by the high current markers contains the RNA-dependent portion of nanopore signal, then that region should differentiate tRNA species. Further, a tRNA-specific change seen in the putative RNA-dependent region, when bordered by the adapter-dependent marker regions, would be evidence that the adapted tRNA translocate entirely through the nanopore. We used the $\phi 29$ DNAP-mediated braking method, as had been done with the RNA(hairpin) substrate, to improve temporal resolution of adapted tRNA. For this experiment we selected two well-characterized tRNA species for nanopore analysis, *E. coli* tRNA^{fMet} and *E. coli* tRNA^{Lys}. These tRNAs exist in the *E. coli* genome as either a single isoform (tRNA^{Lys}) or as two isoforms that differ by only a single nucleotide (tRNA^{fMet}) (RajBhandary 1994). Additionally, tRNA^{Lys} and tRNA^{fMet} have similar lengths (76nt and 77nt respectively), but have significantly different nucleotide compositions (50.0% G-C and 64.9% G-C content respectively) and would be expected to generate different ionic current signals.

Experiments with the adapted tRNA^{fMet} substrate produced 85 events that contained the leading and trailing markers bracketing an extended intervening current region (17.6% of 481 total events) (**Figure 2.5a** and **Figure 2.8**). Experiments with the adapted tRNA^{Lys} substrate produced 77 events that contained the leading and trailing markers also bracketing an extended intervening current region (22.3% of 348 total events) (**Figure 2.5b**). As suggested by the results with the adapted RNA(hairpin) substrate, these events were presumed to result from complete translocation of the adapted tRNA through the nanopore. These events were selected for further analysis of tRNA-specific ionic current signal.

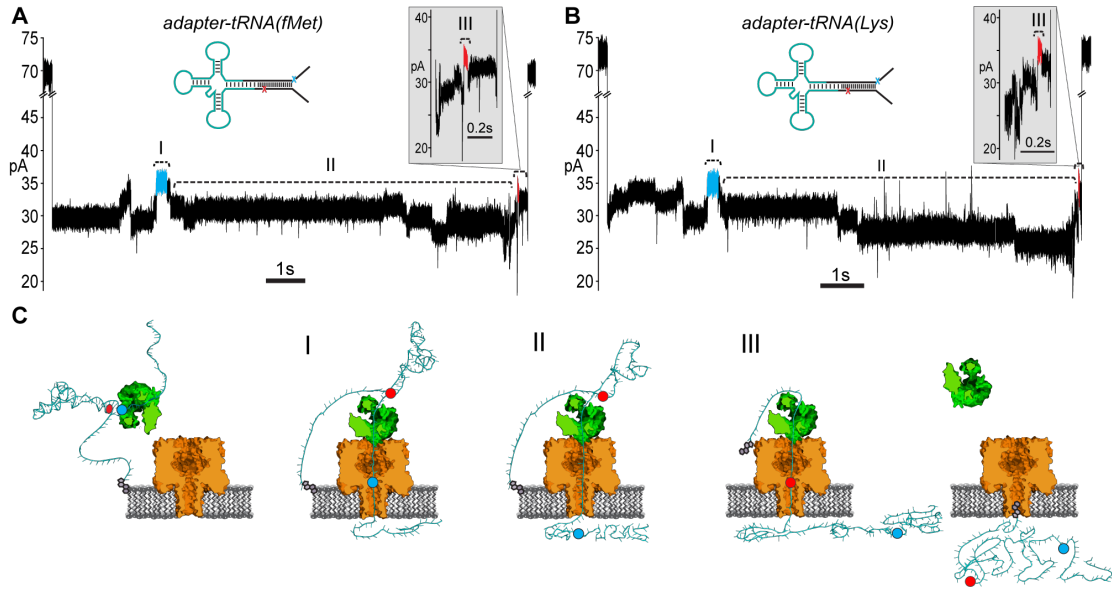


Figure 2.5: Adapted tRNA^{fMet} and tRNA^{Lys} translocate through the α -HL nanopore.

(A) A representative nanopore ionic current trace for adapted *E. coli* tRNA^{fMet}. The rate of translocation through the nanopore is controlled by ϕ 29 DNAP (see Figure 3). The leading high current marker (I, blue segment) and the trailing high current marker (III, red segment) correspond to the 5' abasic residue and 3' abasic residue transiting the nanopore, as in Figure 4B. Region II of the trace contains the portion of the nanopore signal associated with tRNA translocation. Translocation events that contained both 5' and 3' current markers were observed in ~20% of all events with durations exceeding one second. Scale bar indicates one second. The inset shows a magnified view of the trailing 3' high current marker. Scale bar indicates 0.1 second. (B) A representative trace from nanopore experiments with adapted tRNA^{Lys}. Details are the same as in panel A. (C) A cartoon model of an adapted tRNA as it transits the α -HL nanopore. The panels to the left and right correspond to open channel before and after tRNA translocation (see Figure 3). Roman numerals correspond to the roman numerals in panels A and B. I) Translocation of the leading adapter strand through α -HL until the 5' abasic residue (blue circle) reaches the pore limiting aperture. II) Translocation of the tRNA portion of the adapted molecule. III) Translocation of the trailing adapter strand and 3' abasic residue (red circle) through the pore limiting aperture after translocation of the tRNA.

To test the putative RNA-dependent region for tRNA-specific signal(see Fig. 5A-B, *Region II*), we segmented the ionic current signal from each translocation event into regions I through III. As was the case for the RNA(hairpin) substrate (see Figure 4), regions *I* and *III* included the leading and trailing markers respectively, which corresponded to the 5' and 3' abasic residues of the adapter translocating through the nanopore. These served as control regions for our analysis (**Figure 2.5c**). Bracketed by these markers, *Region II* was expected to change based on the identity of the tRNA. Initial inspection of ionic current parameters, mean current and dwell time, suggested that *Region II* provided the best discrimination between tRNA^{fMet} and tRNA^{Lys} (**Figure 2.6**). As expected, mean current and duration for regions I and III did not appear to differ between tRNA^{fMet} and tRNA^{Lys}.

To quantitatively assess the influence of tRNA type on ionic current, we analyzed the data in Figure 6 using a soft-margin support vector machine (SVM) (71). A SVM was used to quantify the discrimination between the two tRNA species using their ionic current parameters in each of the three regions. The SVM produced an optimal linear decision boundary for the event log-durations and mean currents plotted in Figure 6. We used 5-way cross validation to calculate classification accuracy (mean and SD) of the boundaries produced for each of the regions (see Methods). Regions I and III provided discrimination between the two tRNA only slightly better than chance at $60.0 \pm 6.9\%$ and $59.4 \pm 7.0\%$ accuracy, respectively. In contrast, the putative tRNA dependent *Region II* provided a classification accuracy of $87.2 \pm 5.3\%$. This result demonstrated quantitatively that tRNA contributed to the nanopore signal in *Region II*. Further, because the adapter-dependent regions I and III flanked *Region II*, as they do in the adapted tRNA oligonucleotide strand, we concluded that adapted tRNA translocated completely through α -HL as an unfolded, linear strand in these instances.

Discussion

In summary, we have shown that individual biological tRNA molecules can be unfolded and translocated through a nanopore as linear strands. To facilitate this we developed a double-stranded oligonucleotide adapter that could be enzymatically ligated to biological tRNA. The two strands of the adapter acted to locally concentrate adapted tRNA at the bilayer and to initiate strand threading through the nanopore. Using $\phi 29$ DNAP under non-catalytic conditions, we were able to slow strand translocation through the pore. This allowed us to observe adapted tRNA translocating 5'-to-3' through the pore as a linear strand when bound by $\phi 29$ DNAP. Finally, as evidence that tRNA influenced ionic current during translocation we show that *E. coli* tRNA^{Met} and tRNA^{Lys} produced differentiable ionic current signals.

Linear translocation of biological tRNA through the nanopore is a first step towards single-molecule direct sequence analysis of tRNA. Full implementation of a nanopore-based RNA sequencing method will require the following improvements:

- 1) Increased sensitivity of the nanopore. For example, pores derived from mycobacterial species provide nucleotide-level sensitivity for DNA sequencing (11).
- 2) Regulation of tRNA movement in single nucleotide steps. This may be accomplished by coupling an active molecular motor to translocate tRNA, as has been accomplished for DNA sequencing using $\phi 29$ DNAP in a catalytic mode (72).
- 3) Correlation of ionic current states with nucleotide identity. This has been achieved for canonical DNA bases and for various cytosine modifications (11, 64, 73).
- 4) Increased throughput. For example, use of multi-channel nanopore devices will allow for reading tens of thousands of individual tRNA molecules.

Mitochondrial tRNA (mt.tRNA) genes comprise less than 10% of the mitochondrial genome (16,569 bp) yet 43% (245/571) of the currently known pathogenic mutations catalogued in MITOMAP are found in these genes (Brandon et al. 2005, Ruiz-Pesini et al. 2007, Putz et al. 2007). The lack of a direct view of

transcribed mt.tRNA prevents use of sequence, modification state and tRNA abundance as a facile diagnostic in mitochondrial pathology. We expect that a mature nanopore-based method for directly sequencing individual tRNA molecules will yield both canonical base identity and nucleotide modification states along entire strands. This will help advance our understanding of tRNA biology and the diagnosis of mitochondrial pathology as we fully appreciate the role of tRNA in human disease.

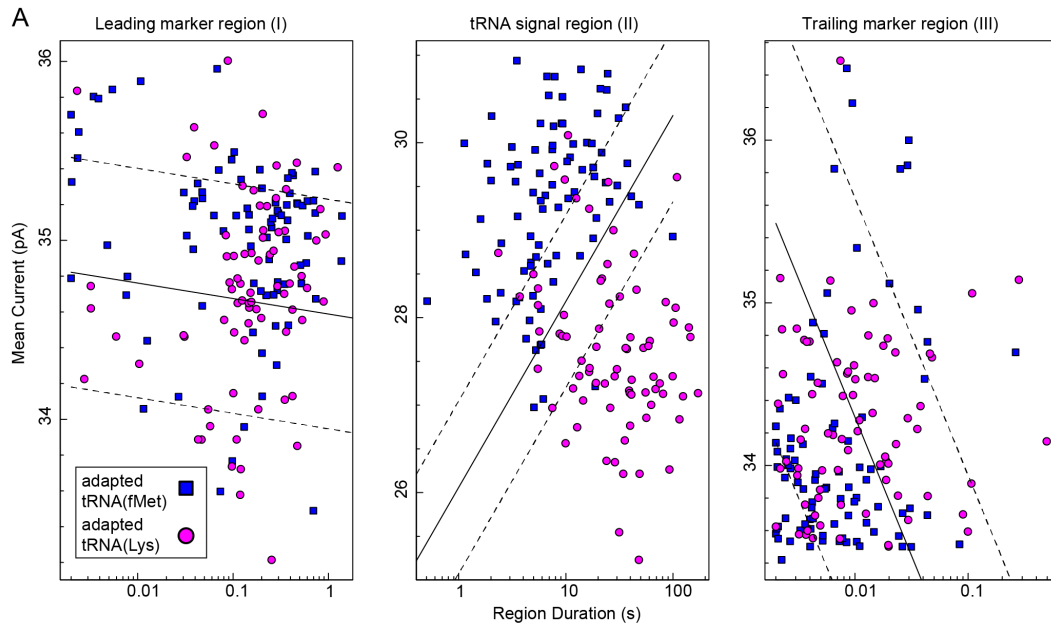


Figure 2.6: Classification of tRNA molecules using duration and mean ionic for regions I-III of adapted tRNA translocation events.

(A) We collected complete ionic current events (approx. 80 for each class) from experiments ($n \geq 5$) using adapted *E. coli* tRNA^{fMet} (blue squares) or adapted tRNA^{Lys} (magenta circles). In our model (Figure 5C), region I and III correspond to the common leading and trailing marker regions, respectively, and region II corresponds to the intervening RNA-dependent portion of the signal. The x-axis is the duration of a given region and the y-axis is mean current for that region. In each panel the solid black line is a semi-logarithmic decision boundary established for that region using a soft-margin Support Vector Machine (SVM) (see Method). SVM margins are shown as dashed lines. The associated classification accuracy for region II was $87.2 \pm 5.3\%$. The associated classification accuracies for region I and III were $60.0 \pm 6.9\%$ and $59.4 \pm 7.0\%$, respectively. The SVM classification accuracy (mean and SD) was established using 5-fold validation (see Methods). Data for the two tRNA species were collected separately from at least five independent nanopore experiments.

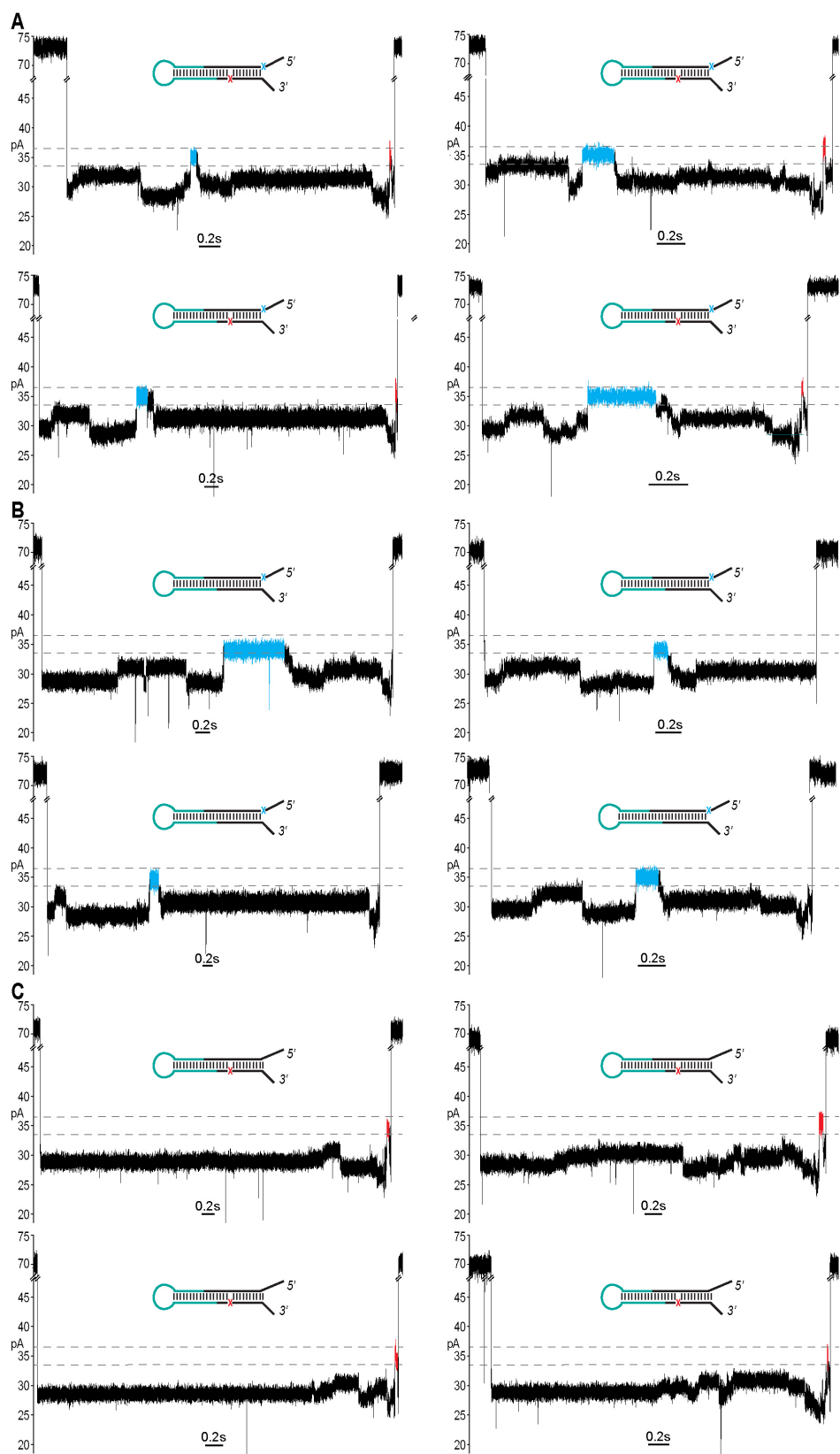


Figure 2.7: Examples of ionic current traces during translocation of RNA hairpins through α -HL with the dual abasic adapter, the 5' mono-abasic adapter or the 3' mono-abasic adapter.

(A) Four examples of ionic current traces for the model RNA (hairpin) containing both the 5' and 3' abasic residues (as shown in Figure 4B). The rate of translocation through the nanopore is controlled by ϕ 29 DNAP (see Figure 3). The events display two high current marker regions characteristic of the 5' abasic residue (blue segment) and 3' abasic residue (red segment) built into the adapter. Scale bar indicates 0.2 seconds. **(B)** Four examples of adapted RNA hairpin containing the 5' abasic residue displaying the leading high current marker (blue segment) (as shown in Figure 4C). Scale bar indicates 0.2 seconds. **(C)** Four examples of adapted RNA hairpin containing the 3' abasic residue displaying the trailing high current marker (red segment) (as shown in Figure 4D). Scale bar indicates 0.2 seconds.

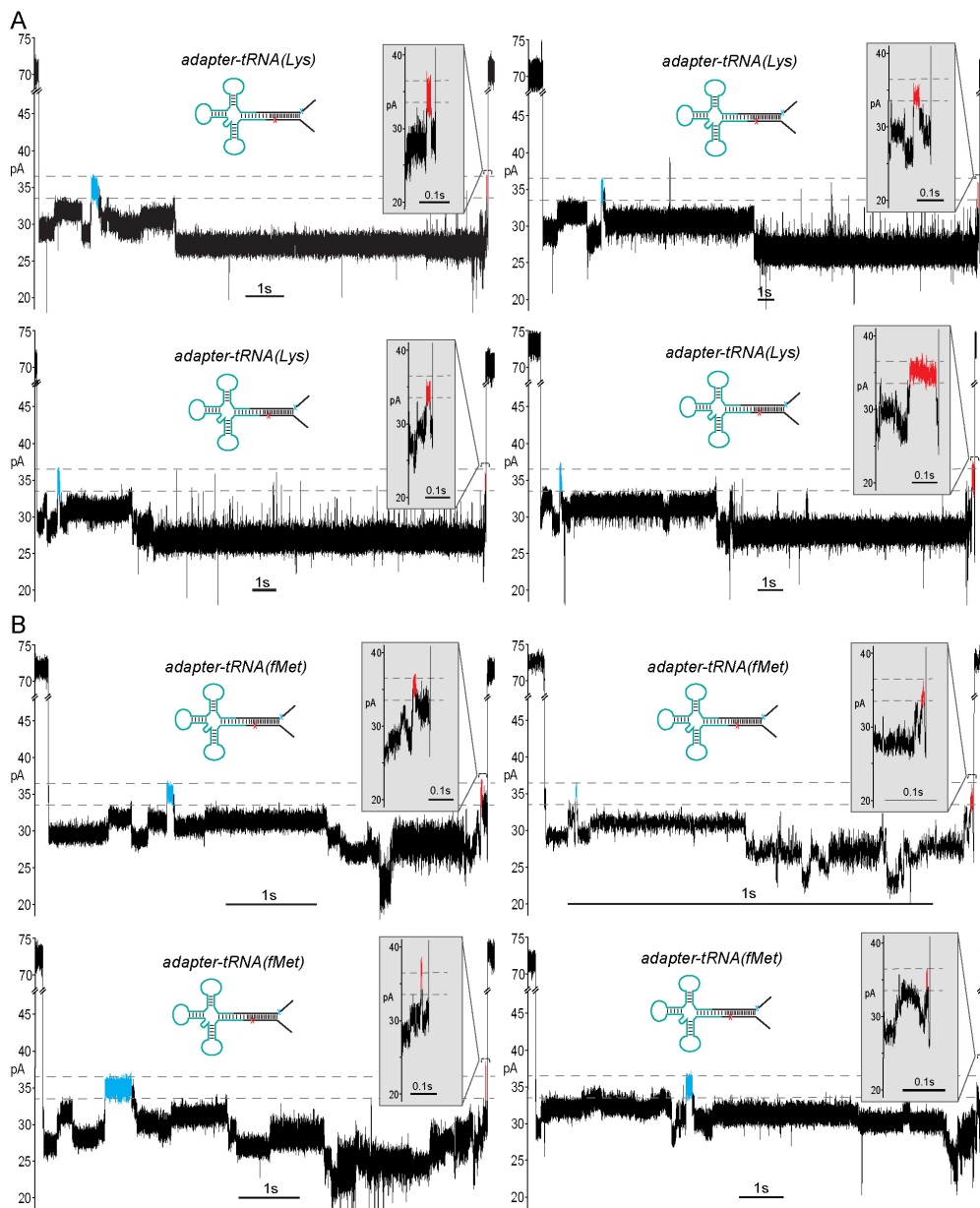


Figure 2.8: Examples of adapted $tRNA^{Lys}$ and $tRNA^{fMet}$ ionic current traces during ϕ 29 DNAP-regulated translocation through α -HL nanopore.

(A) Four ionic current traces during translocation of adapted $tRNA^{Lys}$ molecules are shown. The rate of translocation through the nanopore is controlled by ϕ 29 DNAP. The leading high current marker is highlighted in blue and the trailing high current marker is highlighted in red (as in Figure 5B). Scale bar indicates one second. The inset shows a magnified view of the trailing high current marker. Scale bar indicates 0.1 second. **(B)** Four examples of adapter- $tRNA$ ($fMet$) events are shown. The leading high current marker is highlighted in blue and the trailing high current marker is highlighted in red (as in Figure 5A). Scale bar indicates one

second. The inset shows a magnified view of the trailing high current marker. Scale bar indicates 0.1 second.

Methods

Oligonucleotide synthesis and purification

All oligonucleotides were synthesized by the Stanford Protein and Nucleic Acids facility (PAN) using standard phosphoramidite chemistry. Oligonucleotides were purified by denaturing 7M urea polyacrylamide gel electrophoresis (PAGE) in 1x TBE, followed by overnight elution from an excised band at 4°C in 300mM Sodium Acetate pH 5.2 and 1mM EDTA. DNA was precipitated and recovered by adding 100% molecular biology grade ethanol (Sigma Aldrich) to 70% final v/v and centrifuged for 30 minutes at 14,000 x g and 4°C. Alternately, RNA-containing oligonucleotides were recovered by precipitation in 75% ethanol v/v and centrifuged at 14,000 x g for 30 minutes at 4°C. The ethanol mixture was aspirated and oligonucleotides were then washed with an equal volume of 70% or 75% ethanol and pelleted again for 10 minutes at 14,000 x g and 4°C. Ethanol was aspirated and the pellets were allowed to dry under vacuum to remove residual ethanol. Oligonucleotides were then resuspended in nuclease free water, quantified by Nanodrop (Thermo Scientific), and stored at -80°C.

Adapter design and hybridization

The two oligonucleotides of the tRNA adapter were designed to form a partially double-stranded y-shape. The design included one end that targeted the ACCA tail of tRNA, while the other end contained unpaired single stranded regions (see Figure 2.1a). The adapter strands also included positions where abasic (1'-H deoxyribose) residues were incorporated to act as ionic current markers for the different ends of the adapted RNA molecule.

5' leading strand oligonucleotide (Bolded italicized bases indicate RNA; plain letters are DNA, X indicates an abasic 1'-H deoxyribose): 5'-CTCACCTATCCTTCCACTCATACTATCATTATCTXTCAGATCTCACTA***UCU******GGU***-3'

3' trailing strand oligonucleotide (X indicates an abasic 1'-H deoxyribose; Z is a triethylene glycol-cholesterol): 5'-phosphate-GATXGTGAGATCTGATTTTTTTTTTTTTTTZ-3'

For experiments to determine which end of the adapted RNA molecules entered the pore first, new adapter nucleotides were synthesized with one of the abasic markers (as indicated above) replaced by the nucleotide that restored complementarity in the double stranded portion of the adapter. In these experiments, one of the following oligonucleotides replaced its corresponding abasic-containing strand in the adapter:

5' leading strand oligonucleotide: 5'-CTCACCTATCCTTCCACTCATACTATCATTATCTCTCAGATCTCACTA***UCU******GGU***-3'

3' trailing strand oligonucleotide: 5'-phosphate-GATAGTGAGATCTGATTTTTTTTTTTTTTTZ-3')

Adapters were formed by combining leading and trailing strands at 100μM in 10mM Tris-HCl pH 8 and 50mM NaCl. The mixture was heated to 95°C for thirty seconds and allowed to cool to room temperature.

Nanopore experiments

Nanopore experiments were performed using a single α -HL nanopore (indicated by a step-wise increase in ionic current, range 68.0-72.5pA) embedded in a planar 1,2-diphytanoyl-syn-glycero-phosphatidylcholine bilayer established on a $\sim 25\mu\text{m}$ aperture, as described previously (Akeson et al. 1999). Experiments were conducted in 0.3M KCl and 10mM HEPES pH 8.0 at 28°C ($\pm 0.4^\circ\text{C}$) at 180mV (*trans* well positive). Dithiothreitol (2mM final) and EDTA pH 8.0 (1 mM final) were added to the well on the *cis* side of the bilayer. In the case of experiments looking at the effect of Mg^{2+} , MgCl_2 was added to the buffer to 5mM and EDTA was omitted from the *cis* side well. Nucleic acid substrates were added to 0.5nM unless otherwise noted to the *cis* well and allowed to incubate two minutes to associate with the bilayer. For experiments where $\phi 29$ DNAP was to be added, an additional 12.5 minute incubation was allowed; during this period we observed captures of unbound RNA substrate prior to adding $\phi 29$ DNAP (Enzymatics) to 75nM. Ionic current measurements were collected with an Axon Axopatch 200B (Molecular Devices) patch-clamp amplifier in whole-cell, voltage-clamped mode and filtered with an analog low-pass Bessel filter at 5kHz, then digitized using an Axon Digidata 1440A analog-to-digital converter (Molecular Devices) at 100kHz bandwidth.

RNA, RNA ligations reactions and purification of full-length products

E. coli tRNA^{Lys} and tRNA^{fMet} were purchased from Sigma-Aldrich. The RNA hairpin control was prepared by the Stanford PAN facility using the sequence 5'-phos-CGCGGGGUUUUCCCCGCAACCA-3'. RNA substrates were ligated to adapters using RNA Ligase 2 (NEB). Ligation reactions were carried out in 20 μl of

buffer recommended by the manufacturer supplemented with 0.5mM ATP, 5% PEG 8000, and 2 μ M each of RNA and adapter. To end the reaction and prepare the sample for purification, 50 μ L urea loading buffer (7M urea and 0.1x TBE) was added to the sample and heated to 95°C for 5 minutes. The products were separated on a 7M urea PAGE gel in 1x TBE. The gel was post-stained with 2x SybrGold (Life Technologies) and product of appropriate size for the complete ligation product was excised. Ligated RNA were recovered by electroelution into 3.5kDa MWCO D-tubes (Novagen) at 100V for 2 hours in 1x TBE. Recovered ligated RNA preparations were ethanol precipitated, quantified by Nanodrop (Thermo Scientific), and stored at -80°C.

Event detection and ionic current region measurement

For nanopore experiments that examined populations of ionic current blockade events a custom developed event detection program was used (<https://github.com/jmschrei/PyPore>). The detection algorithm identified ionic current blockades that were self-terminating by selecting for segments that deflected from open nanopore current (68.0-72.5 pA) below a cutoff of 45 or 55pA and with a minimum duration >0.1 millisecond, where voltage was not reversed to eject the strand from the pore (current never <0pA). For experiments where individual ionic current blockade events were examined, raw nanopore ionic current data was filtered with a digital 2kHz low-pass Bessel filter and analyzed using Clampfit 10.4 (Molecular Devices). For ϕ 29 DNAP molecular braking experiments, events were selected from current blockades that had durations greater than one second and self-terminated. Event classification from adapted RNA(hairpin) substrates were analyzed as articulated in the text. For adapted tRNA data, complete translocation events were selected based on the criteria that they contained exactly two abasic-dependent regions; these were defined as high current regions with a mean current greater than

33.5pA, less than 36.5pA, and minimum a duration greater than 2ms. After selection as complete translocation events, duration and mean current for states I-III was measured by hand using Clampfit's internal statistics measuring program.

Semi-logarithmic decision boundary and accuracy derivation

Event durations were transformed to log-durations (\log_{10}) and a linear decision boundary was established using the kernlab(v 0.9-19) package (74) under R (v 3.0.2). The ksvm parameters used were “type='C-svc', kernel = 'vanilladot', C=10” to produce a soft-margin decision boundary. To assess classification accuracy, a five-fold training/test regimen was used. The data set was shuffled and then partitioned into 5 groups of nearly equal size. In a series of 5 tests, one of the 5 groups was withheld as a test set, while the decision boundary was calculated using the remaining 4. This was repeated for each of the 5 groups. This procedure was repeated 50 times, providing 250 assessments of accuracy. Mean and standard deviation of these 250 accuracy scores are reported. For this study, we used a balanced accuracy score, calculated as the mean recall rate for each of the two data classes. For two classes with labels {1, -1},

$$balanced\ accuracy = \frac{\frac{pred(1)}{true(1)} + \frac{pred(-1)}{true(-1)}}{2}$$

where true(n) are counts of test data labeled n, and pred(n) are counts of test data that are correctly classified. The graphic provided in Figure 6 was derived using the full dataset and the kernlab package with parameters as above. Margins (dotted lines) in Figure 6 provide the optimized bounds that maximize the proper classification of labeled data outside the margin while minimizing the cost of misclassified data on the “wrong” side of that margin (71).

Acknowledgements

Oxford Nanopore Technologies provided funding to support this work (SC20130149), α -HL protein, initial study of ϕ 29 DNAP-mediated braking of RNA, and the cholesterol tagging strategy. Jacob Schreiber developed software for nanopore event detection and visualization. Prof. Todd Lowe provided *S. cerevisiae* tRNA^{Phe} and Prof. Harry Noller provided *E. coli* tRNA^{fMet} and tRNA^{Lys}.

Chapter 3

Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing

Introduction

The ribosome small subunit is expressed in all living cells. It performs numerous essential functions during translation, including formation of the initiation complex and proofreading of base-pairs between mRNA codons and tRNA anticodons. The core constituent of the small ribosomal subunit is a ~1.5 kb RNA strand in prokaryotes (16S rRNA) and a homologous ~1.8 kb RNA strand in eukaryotes (18S rRNA). Traditional sequencing-by-synthesis (SBS) of rRNA genes or rRNA cDNA copies has achieved wide use as a ‘molecular chronometer’ for phylogenetic studies (75), and as a tool for identifying infectious organisms in the clinic (76). However, epigenetic modifications on rRNA are erased by SBS methods. Here we describe direct MinION nanopore sequencing of individual, full-length 16S rRNA absent reverse transcription or amplification. As little as 5 picograms (~10 attomole) of *E. coli* 16S rRNA was detected in 4.5 micrograms of total human RNA. Nanopore ionic current traces that deviated from canonical patterns revealed conserved 16S rRNA base modifications, and a 7-methylguanosine modification that confers aminoglycoside resistance to some pathological *E. coli* strains. This direct RNA sequencing technology has promise for rapid identification of microbes in the environment and in patient samples.

Nanopore-based direct RNA strand sequencing (47) is conceptually similar to nanopore DNA sequencing. An applied voltage across a single protein pore in an impermeable membrane results in an ionic current through the pore (3). This current varies when a DNA or RNA strand is captured by the electric field and then moved through the pore in single nucleotide steps regulated by a processive enzyme (9–11). The output is a time series of discrete ionic current segments that correspond to the sequence of bases that occupy the pore at any given time (13, 77).

Other PCR-free RNA sequencing technologies (often referred to as direct RNA sequencing because the RNA is present during sequencing) have been implemented using SBS combined with optical readout of fluorophore-labelled DNA nucleotides (46, 78). They share some of the benefits of nanopore direct RNA sequencing (e.g. absence of PCR biases), however their reported read lengths are short (typically <25 nt and <34 nt respectively).

Direct nanopore RNA sequencing was first implemented by Oxford Nanopore Technologies (ONT) for mRNA using adapters designed to capture polyadenylated RNA strands (47). We reasoned that this technique could be applied to sequence 16S rRNA with modification to the adapter design. 16S rRNA is a logical substrate for nanopore sequencing because of its abundance and broad use for identifying bacteria and archaea. In addition, numerous antibiotics target prokaryotic ribosomes (79), which can acquire resistance via nucleotide substitutions, or by gain or loss of base modifications (31). These base modifications are difficult to detect using indirect SBS methods. A significant advantage of nanopore sequencing is that modifications can be resolved because each nucleoside touches the nanoscale sensor as the strand translocates through the pore.

Results

Figure 3.2a illustrates the strategy we used to prepare 16S rRNA for MinION sequencing. Briefly, 16S rRNA was ligated to an adapter bearing a 20-nt overhang complementary to the 3'-end of the 16S rRNA (**Figure 3.1a** and **Figure 3.2a**). This overhang included the Shine-Dalgarno sequence (80), which targets the conserved anti-Shine-Dalgarno sequence in prokaryotic 16S rRNA (81). Next, a modular Oxford Nanopore Technologies (ONT) adapter bearing a proprietary RNA motor protein was hybridized and ligated to the adapted RNA strands thus facilitating capture and sequencing on the MinION.

Figure 3.2b shows a representative ionic current trace caused by translocation of a purified *E. coli* 16S rRNA strand through a nanopore in the MinION array. The read begins with an ionic current pattern characteristic of the ONT RNA sequencing adapter strand followed by the 16S rRNA adapter strand. The 16S rRNA is then processed through the nanopore one base at a time in the 3' to 5' direction. The ionic current features are typical of long nucleic acid polymers processed through a nanopore (72, 82).

Sequencing of purified 16S rRNA from *E. coli* strain MRE600 produced 219,917 reads over 24 hours that aligned to the reference sequence (16S rRNA *rndD* gene) (**Figure 3.2c**). This represents 94.6% of the total MinION read output for that experiment. Median read length was 1349 bases. We identified 142,295 aligned reads (64.7%) that had sequence coverage within twenty-five nucleotides of the 16S rRNA 5'-end and within fifty nucleotides of the 3'-end.

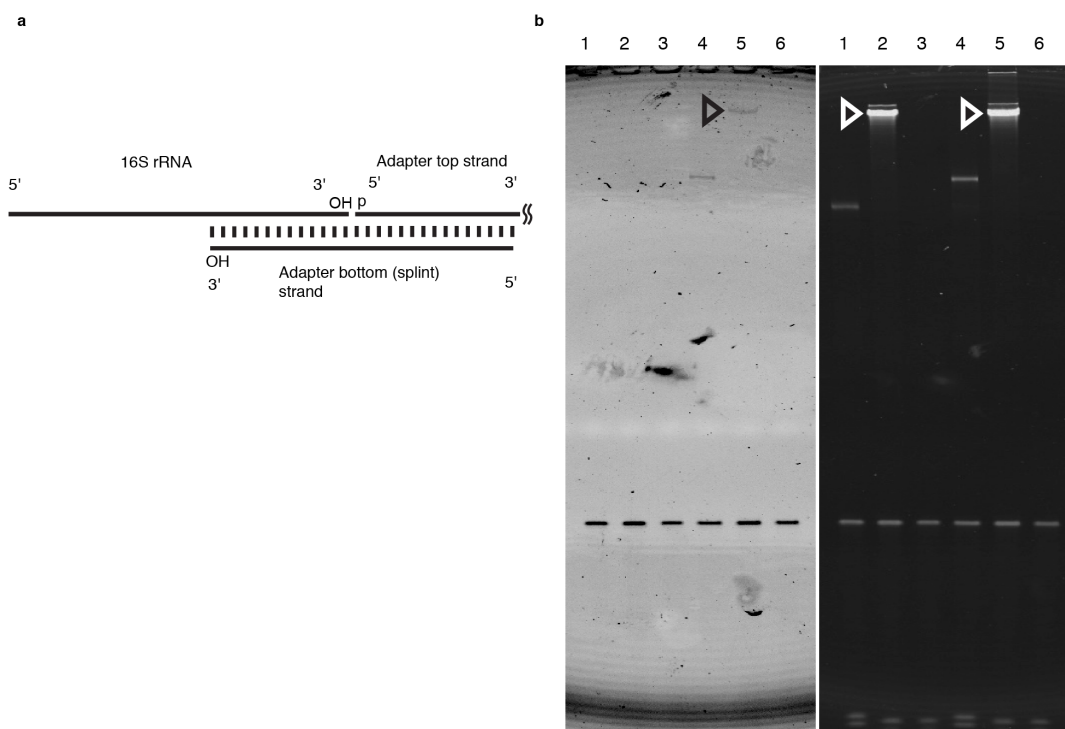


Figure 3.1: Design and testing of an oligonucleotide adapter to prepare 16S rRNA for nanopore direct RNA strand sequencing

(a) Schematic of oligonucleotide adapter hybridized to the 3' end of 16S rRNA. The adapter bottom (splint) strand can hybridize to the conserved anti-Shine-Dalgarno sequence at the 16S rRNA 3' end. The adapter top strand can hybridize to the 5' end of the adapter bottom strand as shown. The 3' terminal hydroxyl of the 16S rRNA and the 5' terminal phosphate of the adapter top strand can be covalently joined by T4 DNA ligase. Overhanging sequence necessary for ligation of ONT direct RNA sequencing adapter is not directly shown. **(b)** Denaturing gel analysis of a ligation reaction demonstrating the 16S adapter hybridizes and ligates to *E. coli* 16S rRNA 3' ends. The left panel shows the unstained gel image. The lower band is a fluorescent, 3' -6-FAM-labeled version of adapter top strand. Lanes 1-3 show pre-ligation reaction samples for: Lane 1) negative control with just adapter present. Lane 2) positive control with an polyA-specific adapter containing a 3' terminal oligo dT 10 overhang (replaces 16S-specific overhang) and the 6-FAM labeled top strand. A synthetic 288mer polyA RNA is used as the control substrate. Lane 3) 6-FAM-labeled 16S rRNA-specific adapter and purified 16S rRNA from *E. coli*. Lanes 4-6 show post-ligation reaction samples for: Lane 4) negative control. Lane 5) positive control with polyA RNA 288mer. Lane 6) 16S rRNA reaction with 6-FAM labeled 16S rRNA-specific adapter. The size-shifted fluorescent top strand indicates ligation to the 16S rRNA 3' end (Open arrow). The right image is the same gel stained with SybrGold. Position of the 16S rRNA is indicated by open arrows.

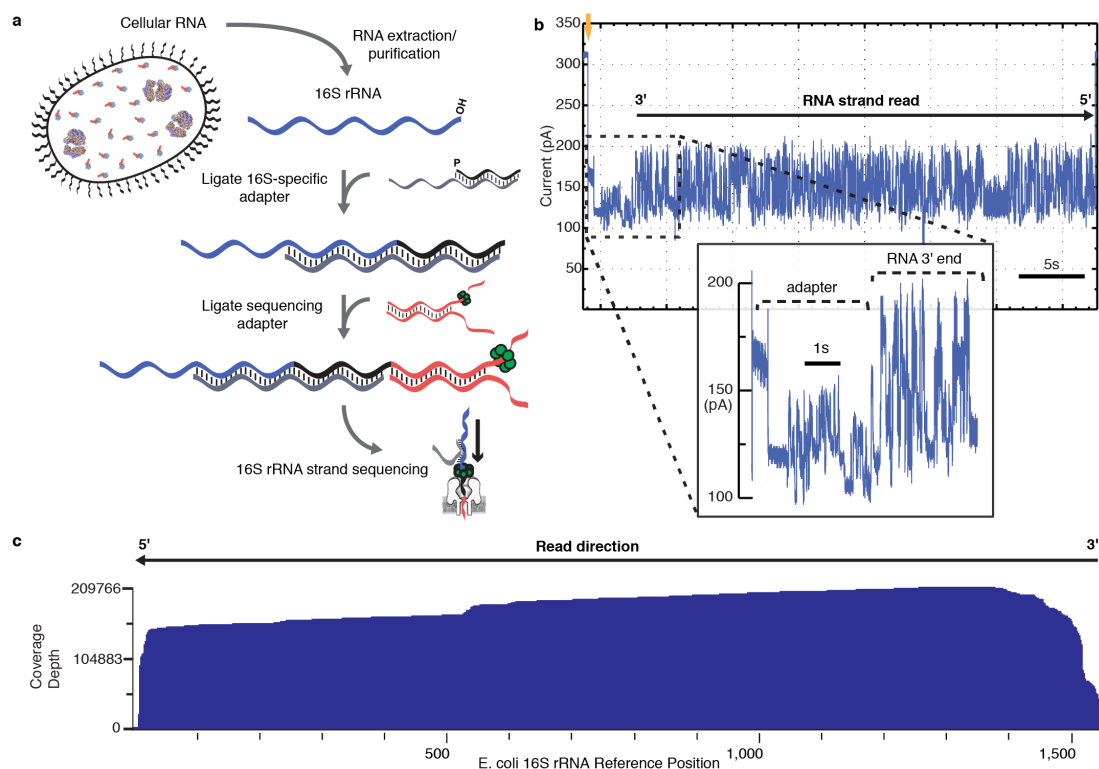


Figure 3.2. Direct nanopore sequencing of individual *E. coli* 16S ribosomal RNA strands.

(a) Library preparation for MinION sequencing. Following RNA extraction, a 16S rRNA-specific adapter is hybridized and ligated to the 16S rRNA 3' end. Next, a sequencing adapter bearing a RNA motor protein is hybridized and ligated to the 3' overhang of the 16S rRNA adapter. The sample is then loaded into the MinION flowcell for sequencing. (b) Representative ionic current trace during translocation of a 16S rRNA strand from *E. coli* str. MRE600 through a nanopore. Upon capture of the 3' end of an adapted 16S rRNA, the ionic current transitions from open channel (310 pA; gold arrow) to a series of discrete segments characteristic of the adapters (inset). This is followed by ionic current segments corresponding to base-by-base translocation of the 16S rRNA. The trace is representative of thousands of reads collected for individual 16S rRNA strands from *E. coli*. (c) Alignment of 200,000+ 16S rRNA reads to *E. coli* str MRE600 16S rRNA *rndD* gene reference sequence. Reads are aligned in 5' to 3' orientation, after being reversed by the base-calling software. Numbering is according to canonical *E. coli* 16S sequence. Coverage across reference is plotted as a smoothed curve. In this experiment, 94.6% of reads that passed quality filters aligned to the reference sequence. Data presented here are from a single flow cell.

We calculated the percent read identities for sequence data from 16S rRNA and Enolase 2 RNA (a calibration standard supplied by ONT) (**Figure 3.3**). The median read identity for 16S rRNA was 81.6% compared to 87.1% for Enolase 2 (**Table 3.1**). Close examination of 16S rRNA reads revealed frequent deletion errors in G-rich regions, which are abundant in non-coding structural RNAs such as 16S rRNA. This is observed as drops in coverage when unsmoothed read coverage is plotted across the *E. coli* 16S rRNA reference (**Figure 3.3d**). Other sequencing errors may represent true single nucleotide variants (SNVs) from the 16S rRNA reference sequence used for alignment. *E. coli* strains typically have seven 16S rRNA gene copies, with some of the gene copies differing by as much as 1.1%. Modified nucleotides could also alter ionic current from canonical nucleotides (48, 83). *E. coli* 16S rRNA contains 11 known nucleotide modifications (30).

We predicted that both SNVs and nucleoside modifications would result in reproducible nanopore base-call errors. Therefore, we looked for positions that were consistently mis-called relative to the *E. coli* MRE600 16S rRNA reference. Using marginCaller at a posterior probability threshold of 0.3 (82), we detected 24 such positions in the nanopore 16S rRNA reads (**Table 3.2**). Five of these were mis-calls resulting from minor variants in the reference sequence relative to the other 16S rRNA gene copies. For example, at position 79 the reference is adenine (A79), whereas the other six 16S rRNA gene copies have a guanosine, in agreement with the majority of nanopore reads. One of the highest probability variants was at G527 in the reference, which was systematically mis-called as a C (**Figure 3.4a & b**). This residue is located in a conserved region of the 16S rRNA 530 loop, near the A-site in the ribosome (84). The guanosine base at this position is known to be methylated at N7 (m7G) (85), which creates a delocalized positive charge. We hypothesized that this modification would significantly alter the ionic current segments that contain m7G527, thus resulting in the systematic base-call error.

Table 3.1: Error rate profile for Enolase 2 transcript and *E. coli* 16S rRNA

Calculated Median	RNA CS (Yeast Enolase 2; 1.35 kb)	16S <i>E. coli</i> rRNA (1.6 kb)
Alignment identity	87.10%	81.59%
Insertions	2.69%	1.90%
Deletions	2.97%	6.02%
Mismatches	5.20%	7.20%
Read coverage	96.14%	97.57%
Read length	1256 bases	1349 bases

Error models were estimated using marginAlign (guide alignments from BWA MEM “-x ont2d” followed by chaining). Statistics were generated using marginStats

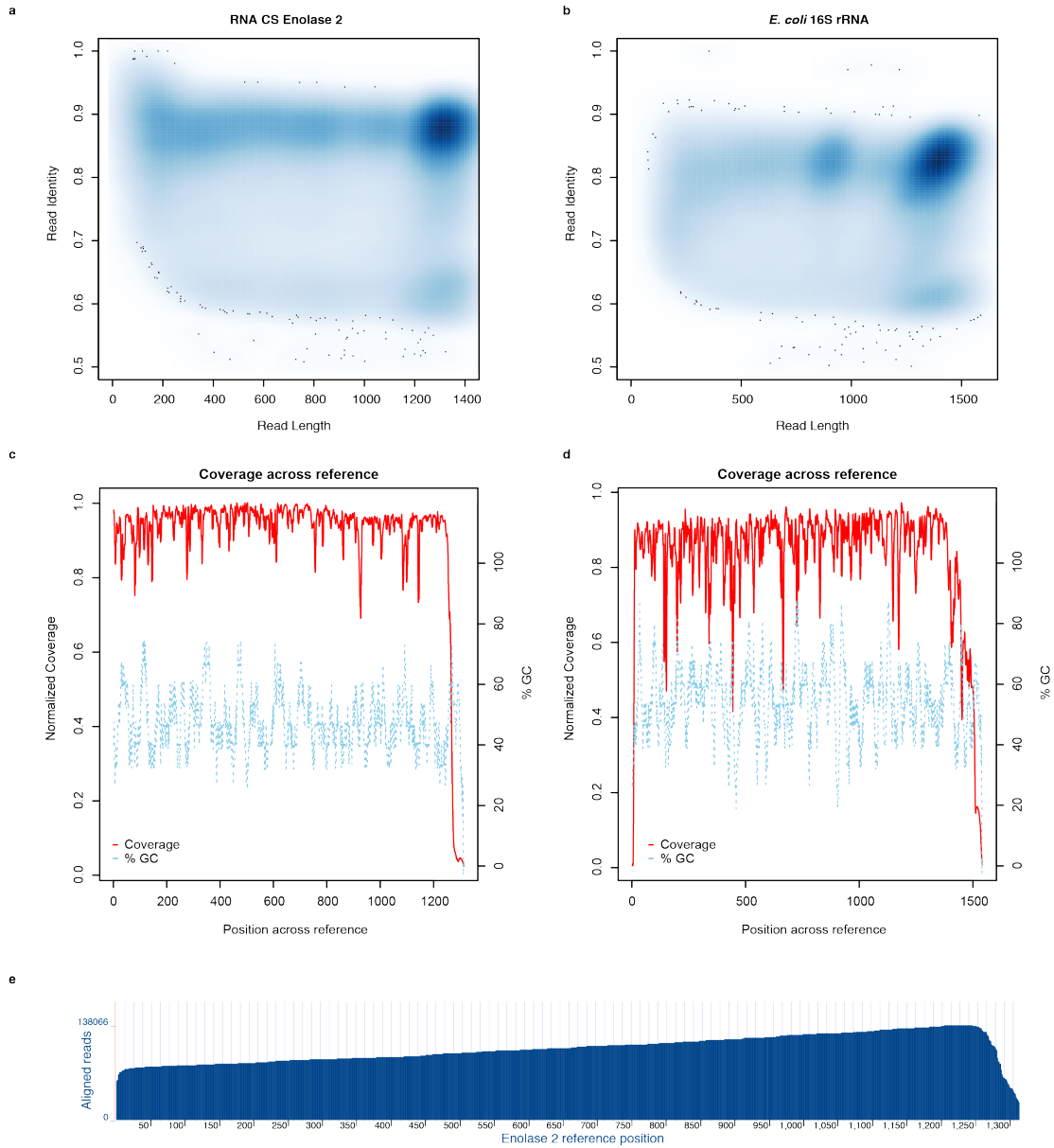


Figure 3.3: Alignment metrics for Enolase 2 polyA calibration strand and *E. coli* 16S rRNA.

MinION sequence data for Yeast Enolase 2 and *E. coli* str. MRE600 16S rRNA were aligned to their respective reference sequences using marginAlign (guide alignments from BWA MEM “-x ont2d” followed by chaining). **(a)** Read identity vs. read length for Enolase 2. **(b)** Read identity vs. read length for 16S *E. coli* rRNA. **(c)** Coverage across reference for Enolase 2 calibration strand. **(d)** Coverage across reference for 16S *E. coli* rRNA. **(e)** Alignment of 100,000+ Enolase 2 reads to the reference sequence.

Table 3.2: Detection of nucleotide variants in MinION *E. coli* 16S rRNA sequencing data using marginCaller

Reference	Position	Ref. base	Var. base	Probability
Ecoli MRE600 rrnD	6	G	A	0.323009035
Ecoli MRE600 rrnD	79	A	G	0.477024303
Ecoli MRE600 rrnD	90	U	C	0.495124202
Ecoli MRE600 rrnD	183	C	U	0.334402802
Ecoli MRE600 rrnD	226	G	A	0.455808427
Ecoli MRE600 rrnD	273	U	A	0.304643919
Ecoli MRE600 rrnD	288	A	G	0.38459441
Ecoli MRE600 rrnD	328	C	U	0.340478205
Ecoli MRE600 rrnD	346	G	A	0.301363107
Ecoli MRE600 rrnD	485	U	C	0.760495972
Ecoli MRE600 rrnD	516*	U	C	0.908639215
Ecoli MRE600 rrnD	527*	G	C	0.778486859
Ecoli MRE600 rrnD	790	A	G	0.496515677
Ecoli MRE600 rrnD	893	C	U	0.43893419
Ecoli MRE600 rrnD	1150	A	U	0.512479558
Ecoli MRE600 rrnD	1195	C	U	0.701573199
Ecoli MRE600 rrnD	1281	U	C	0.702653464
Ecoli MRE600 rrnD	1304	G	A	0.605698867
Ecoli MRE600 rrnD	1380	U	C	0.40811185
Ecoli MRE600 rrnD	1406‡	U	C	0.326475023
Ecoli MRE600 rrnD	1421	G	A	0.317460212
Ecoli MRE600 rrnD	1495‡	U	A	0.568537937
Ecoli MRE600 rrnD	1518*	A	U	0.390200124
Ecoli MRE600 rrnD	1519*	A	U	0.467437353

* modified nucleotide positions in *E. coli*.

‡ positions within 5nt of known RNA modifications in *E. coli*.

Probability indicates a posterior probability that base-call is a true variant from reference as calculated by marginAlign. A cutoff of ≥ 0.3 was used to include only high-confidence variants.

To test this hypothesis, we compared *E. coli* str. MRE600 (wild type) 16S rRNA nanopore reads with reads for an *E. coli* strain that lacks the enzyme RsmG, which is responsible for N7 methylation at G527 (86). We validated the absence of methylation at G527 in the *RsmG* deficient strain by chemical cleavage (**Figure 3.5a**). As predicted, a canonical guanosine base at position 527 in the mutant strain eliminated the reproducible base-call error seen in the wild type *E. coli* strain (**Figure 3.4b**, right panel). Examination of ionic current segments containing G527 and m7G527 in RNA strands for the respective strains confirmed that m7G alters ionic current relative to canonical G (**Figure 3.4c**).

Typically, *E. coli* 16S rRNA contains only one m7G at position 527. However, some pathogenic strains that are resistant to aminoglycoside antibiotics contain an additional m7G at position 1405 (87). The enzymes responsible for G1405 methylation, such as RmtB (88), are thought to have originated from microbes that produce aminoglycosides antibiotics and are shuttled on multidrug-resistance plasmids (89). Given the pronounced signal difference for m7G at position 527, we thought it should also be possible to detect m7G in this context.

To this end, we engineered an *E. coli* strain that carried *RmtB* on an inducible plasmid (pLM1-RmtB, see Methods). We confirmed that this *RmtB*⁺ strain was aminoglycoside resistant, (**Figure 3.5b**) consistent with N7 methylation of G1405. We then compared 16S rRNA sequence reads for this strain (*RmtB*⁺) with reads from the parent *E. coli* strain (BL21) without the plasmid (**Figure 3.4d**). We observed an increase in deletions and base mis-calls in 16S rRNA reads for the *RmtB*⁺ strain at position G1405 and the adjacent U1406. These mis-calls were absent in the 16S rRNA reads for the parent BL21 strain, which bears a canonical guanosine at G1405. Examination of ionic current segments containing G1405 and m7G1405 in RNA strands for the respective strains confirmed that m7G alters ionic current relative to canonical G (**Figure 3.4e**), as was observed at position 527. In this region,

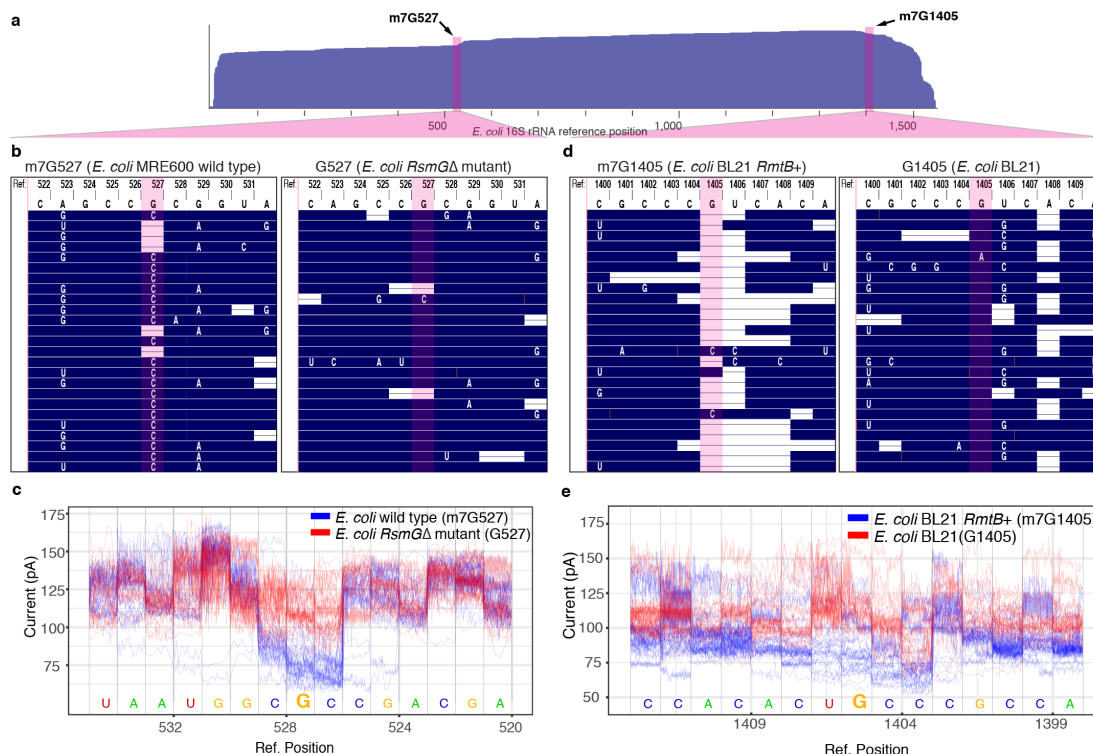


Figure 3.4: Detection of m7G modifications in *E. coli* 16S rRNA.

(a) Diagram showing the positions along *E. coli* 16S rRNA that correspond to the expanded sequence alignments in panels b-e. Arrows indicate the positions of G527 and G1405 in the *E. coli* reference. (b) Alignment of nanopore RNA sequence reads proximal to position 527 of *E. coli* 16S rRNA. Numbered letters at the top represent DNA bases in the reference 16S rRNA gene. Blue regions in the body of the panel denote agreement between reference DNA bases and nanopore RNA strand base-calls. White letters denote base call differences between the reference and the nanopore reads, and horizontal white bars represent base deletions in the nanopore RNA reads. Columns highlighted in red correspond to position 527. The left inset is *E. coli* str. MRE600 (wild type) 16S rRNA (m7G527), and the right inset is *RsmG* mutant strain 16S rRNA (canonical G527). (c) Nanopore ionic current traces proximal to position 527 of the *E. coli* 16S rRNA reference. Blue traces are for wild type *E. coli* 16S rRNA translocation events bearing m7G at position 527. Red traces are for mutant strain 16S rRNA translocation events bearing a canonical G at position 527. The sequence is shown 3'-to-5' with standard *E. coli* numbering right-to-left because ionic current signal is 3'-to-5'. (d) Alignment of nanopore RNA sequence reads proximal to position 1405 of *E. coli* 16S rRNA. Use of colors, shapes, and letters are as described for panel (b). The left inset is engineered mutant *E. coli* str. BL21 (*RmtB*+) 16S rRNA (m7G1405); the right inset is *E. coli* str. BL21 16S rRNA (G1405). (e) Nanopore ionic current traces proximal to position 1405 of the *E. coli* 16S rRNA reference. Blue traces are for mutant strain 16S rRNA translocation events bearing m7G at position 1405. Red traces are for wild type 16S rRNA translocation events bearing a canonical G at position 1405. Numbering is the same as in panel c.

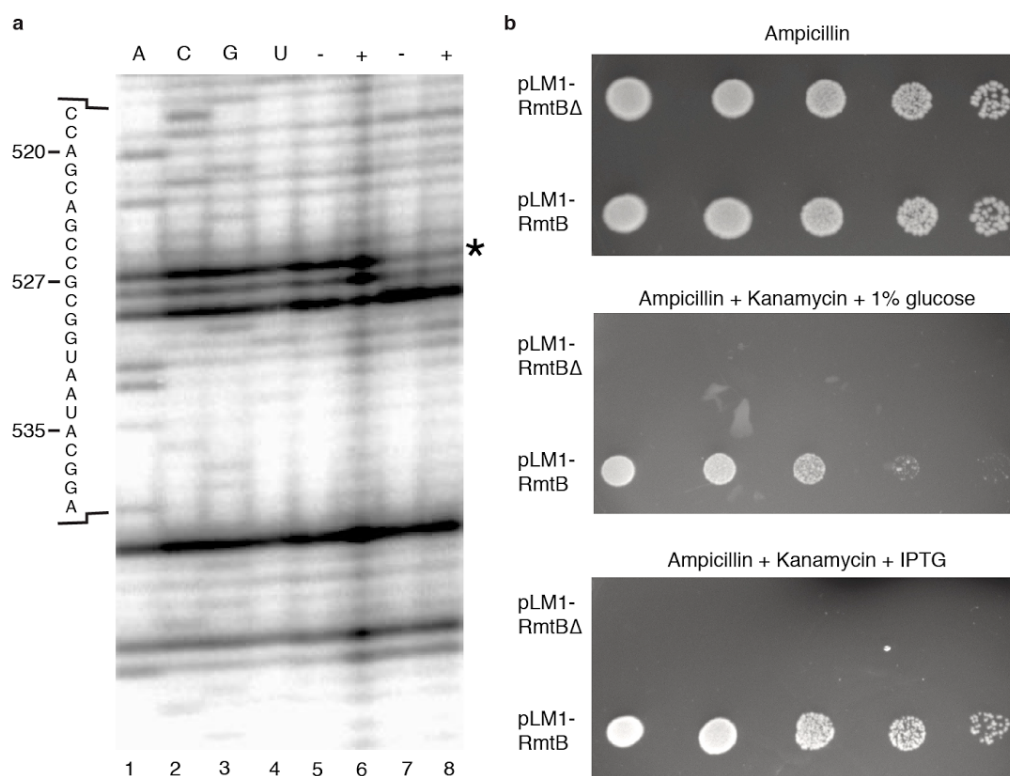


Figure 3.5: Confirmation of m7G at positions 527 and 1405 in *E. coli* 16S rRNA.

(a) Canonical m7G527 is present in wild type *E. coli* and absent in the RsmG deficient *E. coli* strain. Sodium borohydride/aniline cleavage was used to detect N7-methylated guanosine in 16S rRNA for *E. coli* str. MRE600 (wild type) bearing m7G527 and RsmG deficient (mutant) *E. coli* str. BW25113 JW3718Δ. Lanes 1-4 are sequencing lanes for A, C, G, and U respectively. Wild type 16S rRNA from *E. coli* str. MRE600 is used as the template. Lanes 5 and 7: sodium borohydride/aniline treatment (labeled +) of 16S rRNA from wild type *E. coli* and 16S rRNA from RsmG mutant *E. coli*, respectively. Strand cleavage should result in a primer extension stop 1-nt ahead of G527 (position 527 marked by an asterisk). Lane 6 and 8: untreated 16S rRNA for wild type and mutant 16S rRNA. Primer extension products were run on denaturing 6% acrylamide gel, and imaged using a phosphorimager. **(b)** RmtB confers a kanamycin resistance phenotype consistent with G1405 N7-methylation in 16S rRNA from an engineered *E. coli* strain. Serial dilutions from 10^{-2} to 10^{-6} (Left to Right) of *E. coli* BL21 DE3 pLysS strains transformed with pLM1-RmtB and negative control pLM1-RmtBΔ were spotted on LB agar plates. The pLM1 plasmids use pET32a as the backbone, which contains an ampicillin resistance gene. The RmtB gene is under the control of a lactose inducible T7 promoter. Plates are supplemented with: 100 μg/ml Ampicillin (top), 100 μg/ml Ampicillin + 200 μg/ml Kanamycin + 1% glucose (middle), 100 μg/ml Ampicillin + 200 μg/ml Kanamycin + 1 mM IPTG (bottom).

methyated cytosines at positions 1402 and 1407 may also contribute to the aberrant ionic current, which could account for the base mis-calls proximal to those bases in the parent strain (**Figure 3.4d**, right panel and **Table 3.2**).

We then compared 16S rRNA sequence reads for this strain (*RmtB*+) with reads from the parent *E. coli* strain (BL21) without the plasmid (**Figure 3.4d**). We observed an increase in deletions and base mis-calls in 16S rRNA reads for the *RmtB*+ strain at position G1405 and the adjacent U1406. These mis-calls were absent in the 16S rRNA reads for the parent BL21 strain, which bears a canonical guanosine at G1405. Examination of ionic current segments containing G1405 and m7G1405 in RNA strands for the respective strains confirmed that m7G alters ionic current relative to canonical G (**Figure 3.4e**), as was observed at position 527. In this region, methyated cytosines at positions 1402 and 1407 may also contribute to the aberrant ionic current, which could account for the base mis-calls proximal to those bases in the parent strain (**Figure 3.4d**, right panel).

Nanopore detection of epigenetic RNA modifications is not limited to m7G. While examining base mis-calls proximal to G527, we also noted a systematic miscall at U516 (**Figure 3.6a**). This mis-called position had the highest probability variant in our marginCaller analysis (**Table 3.2**). We hypothesized that this was due to pseudouridylation at U516 which is typical in *E. coli* 16S rRNA (90). As a test, we compared nanopore reads for the wild type strain with reads for a mutant strain (*RsuA*Δ) bearing a canonical uridine at position 516. Consistent with the hypothesis, we found that mis-calls and ionic current deviations present at position 516 in the wild type were absent in the mutant strain (**Figure 3.6b**).

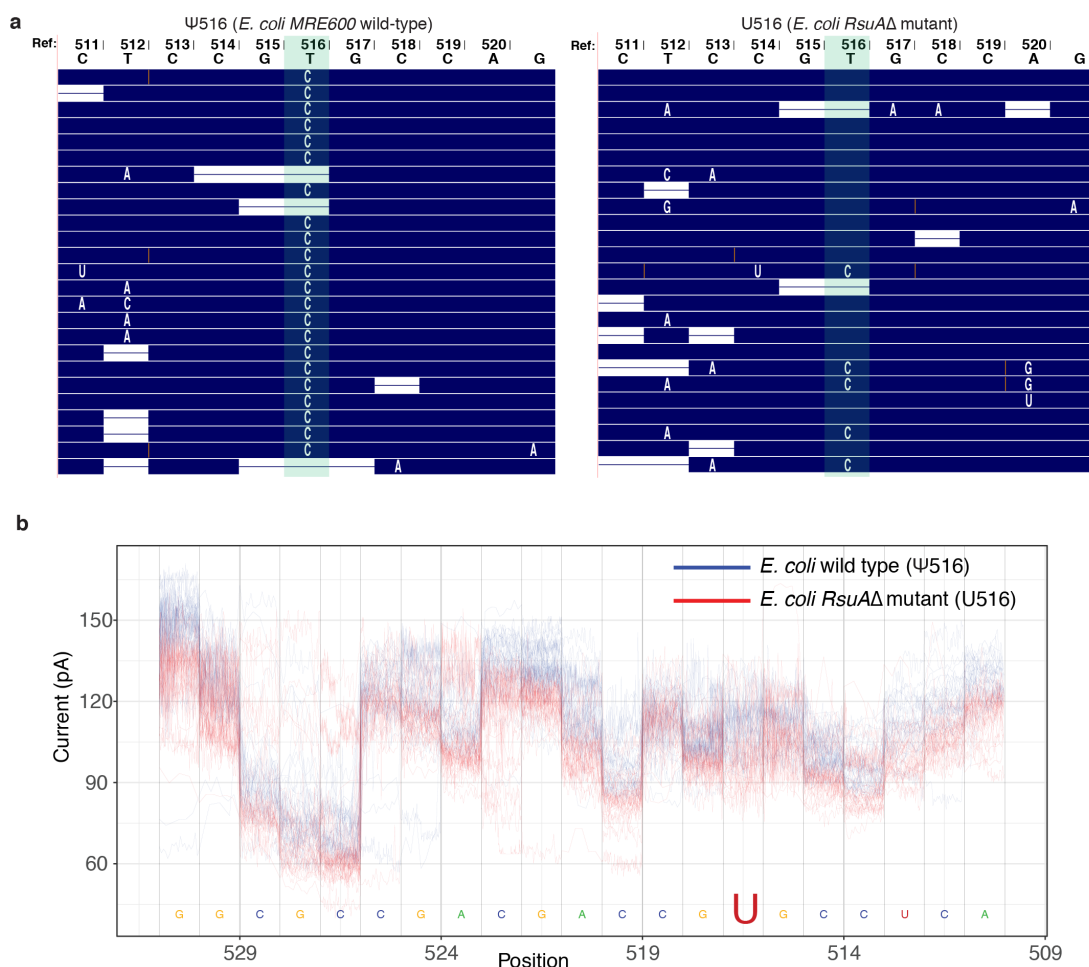


Figure 3.6: Inference of pseudouridine in *E. coli* 16S rRNA direct sequencing reads

(a) Comparison of aligned reads from strands containing putative pseudouridine versus strands bearing canonical uridine at position 516. Reads are aligned to the *E. coli* MRE600 *rnnD* 16S rRNA reference sequence. Shown are twenty-five 16S rRNA reads from separate sequencing runs for *E. coli* str. MRE600 (wild type), which bears a pseudouridine at U516 (Ψ516) and an *RsuA* deficient strain (*RsuAΔ* mutant), which has a canonical U at position 516. Green shading indicates the position of U516 (shown as a T in the reference gene sequence). (b) Aligned ionic current traces from approximately thirty 16S rRNA reads covering position U516 from wild-type *E. coli* and *RsuAΔ* mutant strain. Pseudouridylation site, U516, is shown in large font. The sequence is shown 3'-to-5' because ionic current signal is 3'-to-5'. Numbering uses standard *E. coli* 16S rRNA numbering in right-to-left order.

Another important feature of direct nanopore 16S rRNA reads is that they are predominantly full-length. It has been established that more complete 16S rRNA sequences allow for improved taxonomic classification (91). To test if full-length MinION 16S rRNA reads gave better classification than short reads, we sequenced purified 16S rRNA from three additional microbes (*Methanococcus maripaludis* str. S2, *Vibrio cholerae* str. A1552, and *Salmonella enterica* str. LT2). These were chosen to give a range of 16S rRNA sequence similarities to *E. coli* (68.1%, 90.4%, and 97.0% identity respectively). The 16S rRNA adapter sequence was altered slightly for each microbe (see Methods). We binned reads by length, sampled 10,000 reads per bin for each microorganism, mixed them *in silico*, and aligned them to 16S rRNA sequences for all four microbes. A read was counted as correctly classified if it aligned to a 16S rRNA reference sequence for the source microorganism. As predicted, the classification accuracy increased with read length from 67.9% for short reads (200-600 bases) to 96.9% for long reads (>1000 bases) (**Figure 3.7a**). When using all the reads for each bin per microbe (i.e. no sampling), the average classification accuracy increased to 97.8% for long reads (>1000 bases).

Our early sequencing experiments required purifying 16S rRNA, which is prohibitively slow for clinical applications. Therefore, we devised an enrichment strategy that permits selective preparation of 16S rRNA from total bacterial RNA. This involved adding a desthiobiotin to the 16S rRNA adapter (see Methods). The adapter was hybridized to 16S rRNA in a mixture, and then bound to streptavidin-conjugated magnetic beads. This allowed washing and removal of non-specific RNA. The library preparation was then performed as usual. To test the enrichment method, we prepared 16S rRNA sequencing libraries from the same *E. coli* total RNA preparation with and without the enrichment step. We observed that enrichment increased the number of reads that aligned to 16S *E. coli* rRNA sequence >5-fold relative to the library without enrichment (**Figure 3.7b**).

This suggested that 16S rRNA could be selectively sequenced from a human total RNA background, at relative proportions that would be expected in a clinical sample. To test this, we titrated 5 pg to 500 ng of *E. coli* 16S rRNA into 4.5 µg total RNA from human embryonic kidney cells (HEK 293T) and prepared sequencing libraries (**Figure 3.7c**). The lowest mass (5 pg) approximates the amount of 16S rRNA from 300 *E. coli* cells (92). 4.5 µg of total human RNA approximates the total RNA typically extracted from 1 ml of blood.

We observed a linear correlation between *E. coli* 16S rRNA reads and *E. coli* 16S rRNA concentrations over a 100,000-fold sample range (**Figure 3.7c**). In replicate 5 pg experiments, we observed only 4-5 16S rRNA reads, which nonetheless could be distinguished from the total human RNA negative control (0 16S rRNA reads in 24 hours). Because nanopore data are collected in real-time, we examined how rapidly *E. coli* 16S rRNA was detected in these MinION runs. We extracted acquisition times for all reads that aligned to *E. coli* 16S rRNA (**Figure 3.7d**). At concentrations ≥ 5 ng, we found that the first 16S rRNA read occurred within ~20 seconds of the start of sequencing. This means that some 16S rRNA strands were immediately captured and processed by the MinION upon initiation of the sequencing run. At lower input amounts (<5 ng), we detected *E. coli* 16S rRNA strands in less than one hour. Combined with library preparation, this suggests that nanopore sequencing could detect microbial 16S rRNA in a complex clinical or environmental sample within 2 hours.

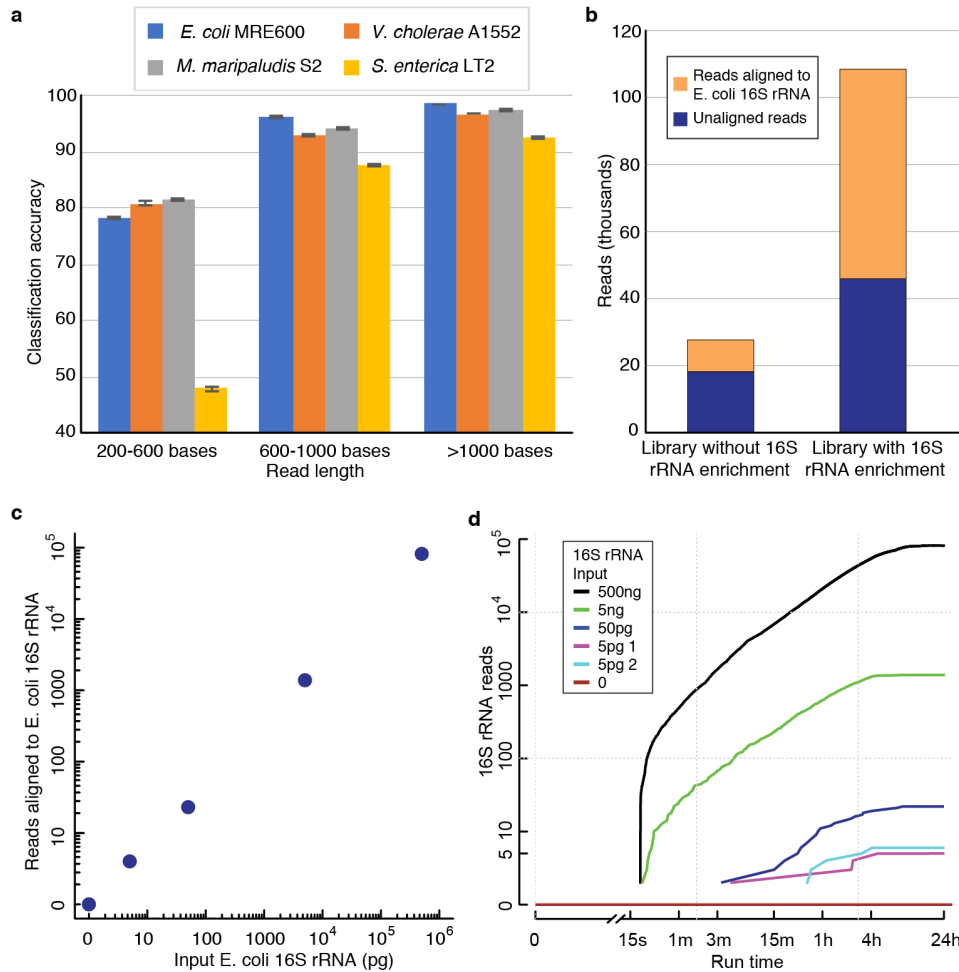


Figure 3.7: Direct 16S rRNA sequencing discriminates among microbes and can detect *E. coli* 16S rRNA at low concentration in a human RNA background.

(a) Classification accuracy from an *in silico* mixture of 16S rRNA reads from four microbes. Reads were binned based on length and 10 iterations of classification using 10,000 randomly sampled reads per microbe were performed. A read was called as correctly classified if it aligned to one of the 16S rRNA reference sequences for that microbe. The error bars indicate one standard deviation for the 10 iterations. (b) 16S rRNA sequencing yield for libraries prepared from *E. coli* str. K12 total RNA with and without enrichment. Sequencing libraries were prepared from 5 μ g total RNA. The enrichment library used a desthiobiotinylated version of the 16S rRNA-specific adapter, which was hybridized and selected for using magnetic streptavidin beads (see Methods). The two 16S rRNA sequencing libraries were then prepared essentially the same way. (c) 16S rRNA reads from sequencing libraries prepared with *E. coli* str. MRE600 16S rRNA titrated into 4.5 μ g total RNA from HEK 293T cells. (d) 16S read accumulation over time in titration sequencing runs. The lines correspond to libraries shown in c.

Conclusion

Some nanopore RNA sequencing applications (e.g. strain-level taxonomic identification or detection of splice sites in transcript isoforms) will require better base-call accuracy than achieved in this study. These improvements seem likely based on prior evidence for MinION DNA sequencing where base call accuracies increased from 66% in 2014 (82) to 92% in 2015 (15).

It is plausible that nanopore RNA sequencing will work for all classes of RNA, with long reads providing more complete sequence and modification annotations.

Methods

Cell culture and total RNA Isolation for 16S rRNA sequencing

E. coli strains BW25113 JW3718 Δ and BW25113 JW2171 Δ (strains hereafter referred to by gene deletion names *RsmG* Δ and *RsuA* Δ , respectively), deficient for 16S rRNA modifying enzymes RsmG and RsuA respectively, were purchased from the Keio Knockout collection (93)(GE Dharmacon). *E. coli* strains K12 MG1655, *RsmG* Δ , *RsuA* Δ and *S. enterica* strain LT2 were grown in LB media (supplemented with 50 μ g/ml kanamycin for *RsmG* Δ and *RsuA* Δ) at 37°C to an A_{600} = 0.8-1.0. Cells were harvested by centrifugation and total RNA was extracted with Trizol (Thermo Fisher) following the manufacturer's recommended protocol. All total RNA samples were treated with DNase I (NEB) (2U/10 μ g RNA) in the manufacturer's recommended buffer at 37°C for 15 minutes. Following the DNase I reaction, RNA was extracted by acid phenol/chloroform extraction (pH 4.4, Fisher Scientific) and two rounds of chloroform extraction. RNA was precipitated with sodium acetate (pH

5.2) and ethanol. RNA was resuspended in nuclease-free water and stored at -80°C. For experiments where human RNA was used as a background, total RNA was extracted from 10⁷ HEK 293T cells following the same steps.

16S rRNA purification

E. coli strain MRE600 16S rRNA was isolated from sucrose-gradient purified 30S subunits. *Vibrio cholerae* strain A1552 and *Methanococcus maripaludis* strain S2 16S rRNAs were isolated by gel purification from total RNA. 50-100 µg total RNA (DNase I treated) was heated to 95°C for 3 minutes in 7M urea/1xTE loading buffer and run on a 4% acrylamide/7M urea/TBE gel for 2.5 hours at 28W. Gel bands corresponding to 16S rRNA were cut from the gel. 16S rRNA was electroeluted into Maxi-size D-tube dialyzers (3.5 kDa MWCO, EMD Millipore) in 1X TBE for 2 hours at 100V. RNA was precipitated with sodium acetate and ethanol overnight at -20°C. RNA was pelleted washed once with 80% ethanol. Recovered RNA was resuspended in nuclease free water and quantitated using a Nanodrop spectrophotometer.

Oligonucleotides and 16S rRNA adapters

The 16S rRNA adapter was designed as a double-stranded DNA oligo. The bottom 40-nt strand has a 20-nucleotide region complementary to the 3' end of the 16S RNA, and another 20-nt region complementary to the top strand (**Figure 3.1a**). This oligo has the sequence 5'-CCTAAGAGCAAGAAGAAGCCTAAGGAGGTGATCCAACCGC-3'. The top strand, which is directly ligated to the 16S rRNA, has the sequence 5'-pGGCTTCTTCTTGCTCTTAGGTAGTAGGTTC-3' (p, 5' phosphate). For *V. cholerae* and *M. maripaludis*, the 3' terminal 20-nt of the bottom strand were slightly

changed to yield adapters with perfectly complementary to their respective 16S rRNA 3' ends. This resulted in the strands 5'-CCTAAGAGCAAGAAGAAGCCTAAGGAGGTGATCCAGCGCC-3' and 5'-CCTAAGAGCAAGAAGAAGCCAGGAGGTGATCCAGCCGCAG-3', respectively. To make a 16S rRNA adapter, top and the bottom strands were hybridized at 10 μ M each in a buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 50 mM NaCl. The mixtures were heated to 75°C for 1 minute before being slowly cooled to room temperature in a thermocycler. We confirmed the adapter hybridizes and ligates to *E. coli* str. MRE600 16S rRNA 3' end by a gel electrophoresis-based assay with a 6-FAM-labeled version of the top strand (**Figure 3.1b**). For experiments where 16S rRNA was enriched from a total RNA background, a desthiobiotin was added to the 5' terminus of the bottom strand. All adapter oligonucleotides were synthesized by IDT.

Purified 16S rRNA Sequencing Library Preparation

Sequencing libraries of purified 16S rRNA for *E. coli* str. MRE600, *V. cholerae* str. A1552, and *M. maripaludis* str. S2 were prepared as follows: 2 pmol 16S rRNA adapter and 1.5 μ g purified 16S rRNA (approximately 3 pmol) were added to a 15 μ L reaction in 1x Quick Ligase buffer with 3000U T4 DNA ligase (New England Biolabs). The reaction was incubated at room temperature for 10 minutes. These reactions were cleaned up using 1.8x volume of RNAClean XP beads (Beckman Coulter), washed once with 80% ethanol and resuspended in 20 μ L nuclease-free water. The RNA sequencing adapter (Oxford Nanopore Technologies) was ligated to the RNA library following manufacturer recommended protocol.

Preparation of RNA Sequencing libraries enriched for 16S rRNA

Enrichment-based 16S rRNA sequencing libraries were prepared for *E. coli* strains K-12 MG1655, BL21 DE3 pLys, BL21 DE3 pLys pLM1-*RmtB*⁺, BL21 DE3 pLys pLM1-*RmtB*^Δ, *RsmG*^Δ, *RsuA*^Δ, and *S. enterica* strain LT2. 16S rRNA-enriched sequencing libraries were essentially prepared as described for purified 16S rRNA with the following exceptions: 15 pmol of 5' desthiobiotinylated 16S rRNA adapter was added to 4.5-5 μg total RNA in 10 μL buffer containing 10 mM Tris-HCl pH 8, 1 mM EDTA and 50 mM NaCl. The mixture was heated to 50°C for 1 minute and slowly cooled to room temperature in a thermocycler (~10 minutes). The mixture was then incubated at room temperature for 20 minutes with 100 μL MyOne C1 magnetic streptavidin beads (Thermo Fisher) in 10 mM Tris-HCl (pH 8), 1 mM EDTA, 500 mM NaCl, and 0.025% NP-40 (Buffer A). The beads were washed once with an equal volume of Buffer A and once with an equal volume of buffer containing 10 mM Tris-HCl (pH 8), 1 mM EDTA, 150 mM NaCl (Buffer B). To elute 16S rRNA-enriched RNA, 20 μl Buffer B amended with 5 mM biotin was incubated with the beads at 37°C for 30 minutes. The hybridized 16S rRNA adapter was then ligated by bringing the mixture to 40 μL 1x Quick Ligase buffer (New England Biolabs) and adding 3000U of T4 DNA ligase (New England Biolabs). The rest of the library preparation was performed the same as described for purified 16S rRNA sequencing libraries.

In vivo methylation of 16S rRNA G1405

The *RmtB* gene was purchased as a synthetic gBlock from IDT with the sequence from GenBank accession EU213261.1. pET-32a+ (EMD Millipore) and

RmtB gBlock were digested with XhoI and NdeI. Digested plasmid and gBlock were ligated with T4 DNA ligase (NEB) to create plasmid pLM1-*RmtB*+. To create *RmtB* null plasmid, pLM1-*RmtB*Δ, XhoI and NdeI digested pET-32a+ was end repaired and ligated. Plasmids were transformed into *E. coli* DH5a cells (NEB) and confirmed by Sanger sequencing. Confirmed clones for pLM1-*RmtB*+ and pLM1-*RmtB*Δ were transformed into *E. coli* BL21 DE3 pLysS cells to create expression strains. To methylate G1405 in 16S rRNA, *E. coli* BL21 DE3 pLys pLM1-*RmtB*+ cells were cultured in 150 ml LB at 37°C with Ampicillin (100 µg/ml) until OD₆₀₀ ~ 0.4. Cultures were diluted into 1 L in pre-warmed LB media with Ampicillin (100 µg/ml), and plasmid expression was induced with 1 mM IPTG. Cultures were grown at 37°C to an OD₆₀₀ ~ 0.4. Cells were then pelleted and resuspended in 30 ml of 25mM Tris-HCl (pH 7.5), 100mM NH₄Cl, 15mM MgCl₂, 5 mM β-mercaptoethanol. Cells were harvested for RNA purification or flash frozen in liquid nitrogen and stored at -80°C.

Chemical probing for m7G

Chemical probing for 7-methylguanosine in *E. coli* 16S rRNA was carried out essentially as described previously (Recht et al. 1996). Approximately 10 pmol 16S rRNA or RNA extracted from 70S ribosomes was resuspended in 20 µl 0.5M Tris-HCl (pH 8.2). Selective reduction of m7G was performed by adding 5 µl freshly made 0.5 M sodium borohydride solution. The reaction was incubated on ice in the dark for 30 minutes. The reaction was ended by the addition of 10 µl 3M sodium acetate (pH 5.2) and precipitated with ethanol. Pellets were washed once with 80% ethanol. RNA was pelleted by centrifugation and resuspended in 20 µl 1M aniline/glacial acetic acid solution (1:1.5) (pH 4.5). RNA cleavage proceeded by incubating the reaction at 60°C for 10 minutes in the dark. The reaction was ended by

the addition of 20 µl 0.5M Tris-HCl (pH 8.2), and the RNA was isolated by extracting with phenol/chloroform/isoamyl alcohol. RNA was precipitated from the aqueous phase, pelleted and washed with 80% ethanol. RNA pellets were resuspended in 2.5 µl nuclease free water. Primer extension to determine the site of m7G-specific cleavage was carried out as described (Merryman and Noller 1998). To detect G527 methylation, the primer 5'-CGTGCGCTTTACGCCCA-3' was used.

MinION sequencing of 16S rRNA

MinION sequencing of 16S rRNA libraries was performed using MinKNOW version 1.1.30. The flow cells used were FLO-MIN106 SpotON version. ONT's Metrichor base-calling software (1D RNA Basecalling for FLO-MIN106 v1.134 workflow) takes this raw signal and produces base-called FASTQ sequence in the 5' to 3' order after reads are reversed. During the course of these experiments, ONT made a new local base-caller available, named Albacore. We performed base-calling for the sequencing runs using Albacore v1.0.1, and performed all alignment-based analyses with the newer sequence data.

MinION data analysis

FastQ sequences were extracted using poretools v0.6 (94) and then sequence alignment was performed using marginAlign v0.1 (82) (using BWA-MEM version 0.7.12-41044; parameter “-x ont2d” (95). The statistics were calculated using marginStats v0.1 (82). We then created assembly hubs to visualize these alignments on the UCSC genome browser using createAssemblyHub utility in marginAlign suite (82). We calculated read identity as matches / (matches + mismatches + insertions + deletions). We used marginAlign expectation maximization (EM) to estimate the

error model from the sequence data. Using these high-quality alignments, we estimated substitution rates for the RNA nucleotides in MinION data. Using these high-quality alignments, we then performed variant calling using marginCaller v0.1 (82) to predict variants and associate systematic sequence mis-calls with putative base modifications. To test for systematic k-mer biases in MinION RNA data, we compared 5-mers in reads and the known 16S rRNA reference.

Nanopore ionic current visualization

We used nanoraw v0.4.2 (96) to visualize ionic current traces for 16S rRNA reads from different *E. coli* strains that were sequenced on the MinION. We used the software with its default settings. We chose graphmap (97) as the aligner in nanoraw, and the argument ‘ont’ (now ‘pA’ in nanoraw v0.4.2) as the option for normalizing raw ionic currents. The ionic current plots were created using the plot_genome_location function. For all of the ionic current analysis, we inverted the reference sequence since the present MinION direct RNA sequencing chemistry sequences native RNA molecules in the 3′-5′ direction.

Microbial classification from 16S rRNA reads

Binning reads by length (200-600, 600-1000, >1000 bases), we randomly sampled 10,000 reads per bin for each microbe. These reads were then mixed *in silico* and aligned using marginAlign v0.1 (82). A read was called as correctly classified if it aligned to one of the 16S rRNA reference sequences for that microbe. 10 classification iterations were performed for each of the bins.

Acknowledgements

Harry Noller and Hossein Amiri provided purified *E. coli* 16S rRNA. Laura Lancaster advised on primer extension experiments. Manny Ares provided assistance with *S. enterica* cell culture and advised on 16S rRNA methylation experiments. David Alexander provided HEK 293T cells. Fitnat Yildiz and Jennifer Teschler provided *V. cholerae* total rRNA. Todd Lowe provided *M. maripaludis* total RNA. Oxford Nanopore Technologies provided direct RNA sequencing kits and Enolase 2 mRNA. Aariah Mackie proofread the final draft. Discussions with James Hadfield reaffirmed our interest in sequencing 16S rRNA. Brad Bebout gave useful feedback on the potential utility of 16S rRNA nanopore sequencing in the field. This work was supported by NHGRI grant HG006321 (MA).

Chapter 4

Active translocation of SV40 Large T-antigen helicase with DNA and RNA substrates atop α -hemolysin nanopore

Introduction

The Large T-antigen (LTag) from Simian Virus 40 (SV40) forms a hexameric, ring-shaped complex at the origin of the SV40 circular chromosome (98). Since discovery of its critical role in unwinding the viral DNA during replication it has been used as a model DNA helicase. Unwinding during replication occurs from two hexamer copies unwinding in opposite directions (99). While various domains of the Large T-antigen have important roles during initiation of DNA replication, the DNA unwinding, or helicase, activity is ATP-dependent and is specific to the C-terminal portion of the complex, which contains a DNA binding domain and the helicase domain (100). LTag was one of the first described DNA helicases (101), and shortly there after it was shown to also have RNA unwinding activity (102). This unwinding activity is 3'-to-5' biased, requiring a 3' single-stranded overhang *in vitro* or the viral replication origin and is stimulated by nucleotide triphosphates (NTPs) (103). Interestingly, the NTPase activity of LTag is biased towards using dTTP and rUTP when dsRNA is the available unwinding substrate (102). In 2004, Xiaojiang Chen and co-workers solved the structure for the helicase domain of LTag (residues 251-627) in presence of ATP, ADP and in the apo state (104) (**Figure 4.1A**). It showed that this

super family 3 of helicase (SF3) shares similar structure to SF3 helicase E1 from bovine papilloma virus and other hexameric helicases (105). It also suggested a mechanism for translocation of DNA through the central channel of the LTag hexamer. Structural rearrangements upon ATP hydrolysis (**Figure 4.1B**) are thought to be transmitted around the ring structure through intersubunit contacts. A recent mutant doping study results suggest that these intersubunit contacts do stimulate NTPase activity in a cooperative, non-random manner, with six active subunits necessary for origin DNA unwinding (106). It remains unclear if LTag subunits hydrolyze and move DNA in a concerted, all-at-once fashion or sequential rotary fashion, as proposed for viral helicase T7 gp4 (107).

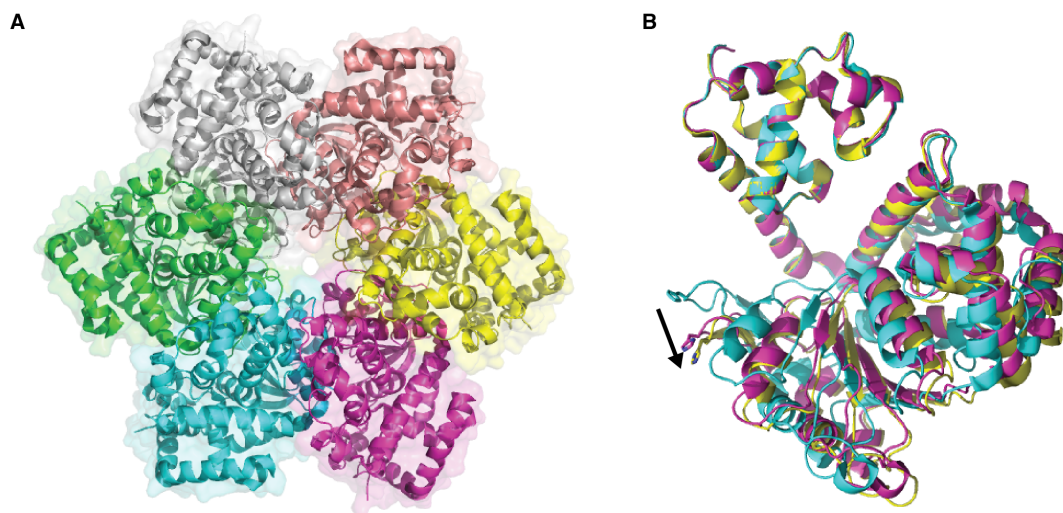


Figure 4.1: Structure of Simian Virus 40 (SV40) Large T-antigen (LTag) helicase fragment

(A) Overhead view of SV40 LTag helicase fragment (residues 251-627) structure model from Gai et al. *Cell*, 2004 (104). The structure shown is in the presence of ATP (PDB 1SVM). The peptide backbone is shown as ribbon model, with surface shading. The central channel is the binding site for nucleic acid strands. Monomer chains are colored separately. (B) Aligned monomer structures in the presence of ATP (cyan), ADP (magenta), apo-state (yellow) as described in Gai et al. Structures were aligned to the N-terminal domain (residues 265-349, upper) to show rearrangements of the lower domain. A conserved histidine is shown in each structure. This histidine is thought to directly contact backbone of the nucleic acid strand being unwound by the helicase. The arrow highlights the relative movement of the histidine side-chain nitrogen at position 3 in the imidazole ring. This position moves ~1nm between the ATP-bound monomer and the apo-state monomer. Structures used were 1SVM, 1SVL, 1SVO, respectively.

This sequential translocation mode is also hinted at by the other SF3 helicase with a solved structure, BPV E1 helicase. It shows conserved histidine residues from five of the six subunits contact the phosphodiester backbone of the substrate DNA strand in the central channel of the helicase in a staircase arrangement (108).

As a helicase that translocates both ssDNA and ssRNA strands, the LTag helicase is an attractive target for use as a molecular motor to control RNA movement through a nanopore. It would be desirable to move RNA through the pore in single nucleotide increments, which is assumed to be a prerequisite for being able to sequence RNA using a nanopore. The additionally, if LTag helicase contacts the phosphate backbone in the same manner as BPV E1 helicase, it should be able to accommodate bulky modified ribonucleotides without steric interference with the helicase. This would facilitate sequencing of native RNA strands with modified nucleotides.

To test the potential for LTag helicase as a RNA motor atop a nanopore we prepared active an hexameric LTag helicase fragment (residues 262-627) as described by Gai et al 2004 (109). This helicase could be produced in a quantity by a bacterial expression system and helicase had NTPase and DNA/RNA translocase activity. Using a well-established, single-channel nanopore setup (5), we demonstrate that actively translocating LTag/nucleic acid strand complexes could be observed, which prolong the dwell time of DNA and RNA/DNA chimeric strands in the pore. We exploit apparent pausing of the helicase when transitioning to a ribonucleotide strand to examine ionic current amplitudes correlated with single nucleotide advancement of these reporter strands. These results imply single nucleotide advancement of the LTag helicase fragment on these substrates do occur while captured in the nanopore. However, it is noted that the putative single nucleotide displacements do not appear to be the only translocation step of the LTag helicase fragment, since amplitudes associated with single nucleotide displacement were not observed in every complex. Additionally, these results do not necessarily imply that the helicase functions in a

biological setting in fundamental single-nucleotide steps, nor can the possibility of non-biological LTag helicase complexes being captured and examined here be ignored.

Results

We set about testing the LTag helicase as a nanopore motor by selecting a helicase construct that has been previously demonstrated to form active, soluble hexamers when overexpressed and purified from *E. coli*. The LTag helicase fragment (residues 262-627) described by Gai et al. was selected based on reported results for various LTag constructs (109). The protein was expressed as a GST-fusion protein, and the GST domain was removed proteolytically after the initial affinity purification. Stable hexamer assembly occurs *in vitro* in buffer containing ATP (109), and a UV absorbance peak of approximately the correct size (240kDa) was observed during preparative size-exclusion chromatography (**Figure 4.2A**). Fractions corresponding to this peak were collected. Analysis by SDS-PAGE showed a highly pure protein sample with a single band of appropriate size for a monomer of the 262-627 LTag helicase construct (here after referred to as LTag helicase) (**Figure 4.2B**).

The activity of the putative purified LTag hexamers was tested with two biochemical assays. The first assay was a colorimetric assay for phosphate. An active NTPase, like LTag, should produce detectable levels of phosphate as a byproduct of NTP hydrolysis under active conditions. The results of the assay show that LTag hexamers were active NTPases (**Figure 4.2C**). We determined that LTag hexamers were active in a range of KCl concentrations that compatible with nanopore analysis. Notably the NTPase activity was higher at 37°C than at 30°C, but activity was still detectable. We also noted that NTPase activity was stimulated by the addition of single-stranded RNA, which is consistent with the literature on active full-length LTag hexamers (110).

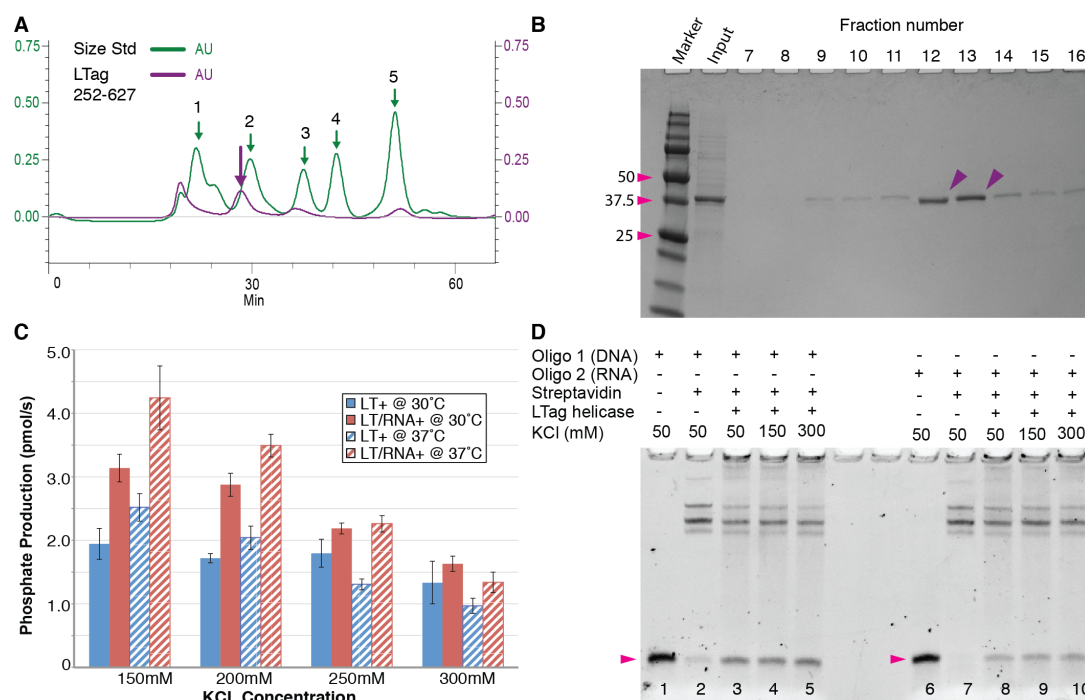


Figure 4.2: Formation of activity helicase complexes by construct LTag 262-627 under buffer conditions compatible with nanopore analysis

(A) Chromatogram from size exclusion purification of putative hexamers formed by the LTag helicase construct described in Gai et al. Cell, 2004 (residues 262-627, hereafter referred to as LTag helicase). After expression and purification as GST-fusion protein in *E. coli*, the LTag helicase was subjected to size-exclusion purification to isolate putative hexameric helicases. During the LTag sample run (purple trace) fractions were collected surrounding the putative LTag hexamer peak (purple arrow). Green trace shows a size standard run on the same column (Superdex 200 10/300) in the same buffer. Peak numbers correspond to: 1) Thyroglobulin (660 kDa, homodimer), 2) γ -globulin (158kDa), 3) ovalbumin (44kDa), 4) myoglobin (17kDa), 5) vitamin B (1.3kDa). (B) Fractions collected from size exclusion chromatography of LTag helicase analyzed by SDS-PAGE. Left-to-right, the gel shows a size standard (Thermo-Fisher #22610), SD200 input from the LTag helicase prep, and eluted SD200 column fractions. Magenta arrows indicate the sizes of relevant standards in kDa. The putative LTag helicase hexamer peak (see A) is eluted in fractions 12-13. These were pooled and concentrated and used in subsequent experiments. The putative monomer LTag (42kDa) is indicated in these fractions by a magenta arrow. (C) NTPase activity of purified LTag helicase in buffers containing varying levels of potassium chloride. Hydrolysis rates for 7.5pmol LTag helicase hexamer was measured by colorimetric malachite green assay (see methods) in the presence of 1mM dTTP. Assays were performed in triplicate at 30°C and 37°C. Stimulation of RNA (polyU) on NTPase activity was also measured. Error bars represent one standard deviation of the mean. (D) Native PAGE analysis of a single-stranded translocation assay for LTag helicase translocation activity. The assay measures helicase translocation by displacement of streptavidin from complexes formed with 5'-biotinylated single-stranded DNA or DNA/RNA chimeras (see Morris et al., Biochemistry 2002). Oligo1 is DNA 46mer, and Oligo2 is a RNA/DNA chimera (15nt RNA, 31nt DNA). The positions of oligos not complexed with streptavidin are indicated on the gel by a

magenta arrowhead. Streptavidin has four biotin binding sites. After complexes between the oligos and streptavidin are formed excess free biotin is added to the reactions (see methods). dTTP is present in the reaction at 1mM. KCl concentration in buffer is indicated above the lanes.

The second assay was a single-stranded nucleic acid translocation assay in which full-length LTag hexamers have been demonstrated to work (111). This assay used synthetic 46mer oligonucleotides with a 5'-biotin. Streptavidin is pre-bound to these oligos. Active 3'-to-5' helicases, such as LTag, then can displace bound streptavidin from the oligo under conditions that are supportive for helicase activity. Displacement of streptavidin, and therefore strand translocation activity, are observed by oligos that migrate more rapidly than streptavidin-bound oligos in non-denaturing PAGE analysis. Single-strand translocation activity was observed with an all DNA oligo, and, at a reduced level on a RNA/DNA chimeric strand (**Figure 4.2D**). Little detectable strand translocation activity observed with oligos composed of RNA only (data not shown). These results indicate that the purified protein was in fact an active form of the LTag helicase and, at least upon initial purification, a hexamer.

Long ionic current blockades indicate active LTag/DNA complexes can be captured atop a nanopore

Having determined that the purified LTag complex formed hexameric complexes capable of hydrolyzing NTPs and translocating along single nucleic acid strands, we set about testing if complexes between the LTag helicase construct and a DNA substrate could be captured on a nanopore. To do this we designed a self-annealing hairpin design with a 14 base-pair duplex and a short 4 nt loop (**Figure 4.3A**). The substrate also included a 35 nt 3' single-stranded portion that was intended to act as a binding site for LTag and to initiate threading of the hairpin through the nanopore (66). Experiments were performed with α -hemolysin (aHL) nanopore in buffer conditions that support LTag activity in our *in vitro* assays (see Methods).

When the hairpin was added to the *cis* well of the nanopore set up on its own only short duration ionic current blockades (10-100ms) were observed (not shown). This is consistent for a short DNA hairpin substrate translocating through the pore. However, once LTag was added to the *cis* well in the presence of rATP or dTTP and MgCl₂, ionic current blockades with durations in excess of 0.1 second were observed. Many blockade events persisted over 1 second. Inspection of these events showed reproducible patterns in the ionic current, typically with amplitudes starting around 40 pA and staying in the range of 20-40 pA. In 8.1% of blockade events (31 of 384) with a duration ≥ 0.2 seconds, a distinct square shaped step up to a maximum amplitude of ~43 pA was observed (**Figure 4.3B** and **Table 4.1**). This high amplitude feature likely arises from the 5-abasic residue (1', 2' H) block in the 35 nt ssDNA 3' tail of the hairpin substrate passing through the pore (**Figure 4.3B, lower**). The shape of the ionic current pattern is consistent with the pattern produced by a 5-abasic block passing through aHL nanopore when non-catalytic phi29 DNA polymerase is controlling moving of a DNA strand in a nanopore (9). It is notable that amplitudes of the putative abasic dependent ionic current are higher here, but the well temperature is elevated relative to conditions that phi29 DNA polymerase has been used with a nanopore.

We hypothesized that these long duration ionic current blockades represented catalytically active complexes of LTag and the DNA hairpin substrate. If this is true, translocation of DNA should correlate with NTP hydrolysis by LTag. To test this hypothesis, we spiked in a non-hydrolyzable ATP analog, γ -S-ATP, to stall NTP hydrolysis by LTag subunits. The prediction being that the ionic current blockades should be extended in their duration. Upon addition of 0.33mM γ -S-ATP (1:6 ratio with dTTP) ionic current blockades were clearly extended. When concentrations of γ -S-ATP were increased (1mM) some blockade events persisted for hundreds of seconds (not shown). Notably, many events had a similar series of mean amplitudes,

suggesting the same DNA sequence was passing through the pore, but slowly. This is consistent with the hypothesis that these blockades arose from LTag actively translocating through the DNA hairpin, with the 5-abasic block acting as a reporter of the passage of the DNA through the pore. Importantly, no long blockades were produced with LTag present in the *cis* well of the nanopore apparatus when the DNA hairpin was not present. This rules out the possibility that the long ionic current blockades arise from LTag interacting with the pore on its own.

Ionic current blockades from LTag complexed with chimeric RNA hairpins indicate RNA induces helicase pausing

With the previous result suggesting that active LTag/DNA complexes were captured and analyzed in aHL nanopore, we sought to demonstrate active control of RNA strands in a nanopore. Given that no or little translocation activity was observed with pure RNA substrates in the translocation assay we designed an RNA/DNA chimeric hairpin (#2028) that maintained the same overall sequence and length as our DNA substrate, with the exception that the hairpin was composed of RNA (**Figure 4.4A**).

This substrate in the presence LTag produced long ionic current blockades similar to the DNA substrate (**Figure 4.4B**). However, a larger proportion of events (~43%) contained a series of high amplitude (>40 pA) states (**Table 4.1**). Notable among this fraction of blockade events were ones where the ionic current fluctuated between two or three distinct amplitudes (**Figure 4.4C**). These events also typically had longer overall dwell times. These fluctuations were reminiscent of angstrom-scale movements observed with DNA being drawn through the pore by the catalytic cycle of phi29 DNAP (10, 112, 113). Given the previously established sensitivity of aHL nanopore to small movements of a 5-abasic reporter, we hypothesized that these fluctuations may represent stalls of the active LTag helicase when it transitioned from






binding DNA to RNA, and these stalls are being reported out as the 5-abasic block just enters the sensitive region of aHL nanopore (114). This hypothesis would predict that replacing ribonucleotides with corresponding deoxyribonucleotides should eliminate the stalling, resulting a reduction in the number of LTag-bound events that display the high amplitude states associated with the 5-abasic reporter block.

To test this prediction we made a series of RNA bearing hairpins. These hairpins placed RNA on either side of the duplex while the remaining nucleotides were replaced with DNA (substrate #2031A and #2032A). The sequence and design of the hairpins, including the 5-abasic reporter remained otherwise unchanged. As shown in **Table 4.1**, the hairpins that contained RNA immediately following the single-stranded 3' tail resulted in a large fraction (37-43%) of blockade events that bear the high amplitude reporter caused by the 5-abasic block. This is consistent with the RNA-induced helicase pausing hypothesis. Notably, the substrate that bore RNA only on the 5' end of the hairpin produced long blockades as well, but very few presented the high amplitude marker. This is also consistent with the RNA-induced helicase pausing hypothesis and can be interpreted that most of these LTag complexes had succeeded in translocating into the chimeric DNA/RNA duplex when captured in the pore.

An alternative hypothesis to RNA-induced pausing is that the LTag helicase struggles to unwind A-form duplex that is typically adopted by RNA and DNA/RNA chimeras in solution (115). To test this alternative we designed an additional RNA hairpin substrate (#2033). This substrate contained six unpaired abasic residues (1',2' H), which should aid in unzipping the 14 bp RNA duplex. This substrate gave similar numbers of events to other substrates bearing RNA at the base of the single-stranded region (**Table 4.1**). This result supports the original RNA-specific induced pauses of

the LTag helicase, and is not consistent with the alternative explanation that simple duplex structure is sufficient to pause LTag helicase fragment on a strand.

Table 4.1: Quantification of active LTag/hairpin complexes captured on a nanopore before the 5-abasic block is translocated through the pore

Design (magenta = RNA, black = DNA)	Hairpin	Putative Binding Events ^a	Abasic marker events ^b (start before 40pA segment)	Total % of events with abasic marker
	DNA hairpin (substrate #2025)	384	31	8.1%
	RNA hairpin (#2028)	308	132	43%
	RNA/DNA hairpin (#2031A)*	332	124	37%
	DNA/RNA hairpin (#2033)*	341	39	11%
	RNA hairpin (unpaired 5' end) (#2032A)*	270	111	41%

*Substrates were designed to test hypothesized RNA-induced pausing of LTag helicase

^a These are defined as any ionic current blockade longer than 0.2 seconds

^b These binding events start with an amplitude less than 40pA and then traverse through a state ≥ 40 pA before returning to amplitudes typically associated with canonical nucleotides translocating through α HL nanopore.

Amplitudes associated with RNA-induced pauses are specific to the location of the abasic reporter

The results from the series of RNA hairpin substrates and bulk translocation assays suggest that RNA on a DNA/RNA chimera may retard LTag helicase translocation. With this in mind, we sought to correlate the abasic reporter position in

the pore with these putative RNA-induced pauses of the LTag helicase. To do this we designed a two other RNA hairpin substrates, which moved the 5-abasic block ± 1 nt from the base of RNA duplex (#2029 and #2030) (**Figure 4.5A**). These substrates should position the 5-abasic block 1 nt closer or further from the sensitive part of aHL nanopore at the time that LTag contacts RNA of the duplex (**Figure 4.5B**, right). Numbering relative to the first nucleotide of the RNA duplex would place these abasic residues at -11 through -15 for #2030 or -13 through -17 for #2029.

Nanopore experiments with the substrate #2030 and catalytic LTag resulted in reproducible events where the prolonged dominant amplitude shifts lower to ~ 28 pA (**Figure 4.5B**, upper). On the other hand, nanopore experiments with substrate #2029 and LTag shifts the dominant amplitude in these events upwards to ~ 40 pA (**Figure 4.5B**, lower). The original RNA hairpin substrate often produced events with a dominant amplitude in between the two other RNA hairpin substrates **Figure 4.5B** (middle). Putting these three amplitudes together in a series reconstructs ionic current amplitudes associated with 3 single-nucleotide increment displacements of a strand through the aHL nanopore. This series of amplitudes was often observed in both RNA and DNA hairpin substrates while active LTag is presumed to be translocating. Moreover, this result is consistent with bound LTag helicase pausing at the same position on all three RNA hairpin substrates (#2028, 2029 and 2030), with the change in amplitudes arising only from moving the location of the abasic report in the design of the substrate.

The implication of the above result is that the three amplitudes steps observed in ionic current blockades correspond to single nucleotide steps of the active LTag helicase at the initial unwinding of hairpin duplex. These states are illustrated for both the original DNA hairpin substrate (#2025) and the original RNA substrate (#2028) in **Figure 4.6A** and **B**. The amplitudes and frequency of these steps were analyzed by hand-segmenting events for both DNA and RNA hairpin substrates that contained the full ionic current pattern (**Figure 4.6C** and **D**). The amplitudes generally agree for

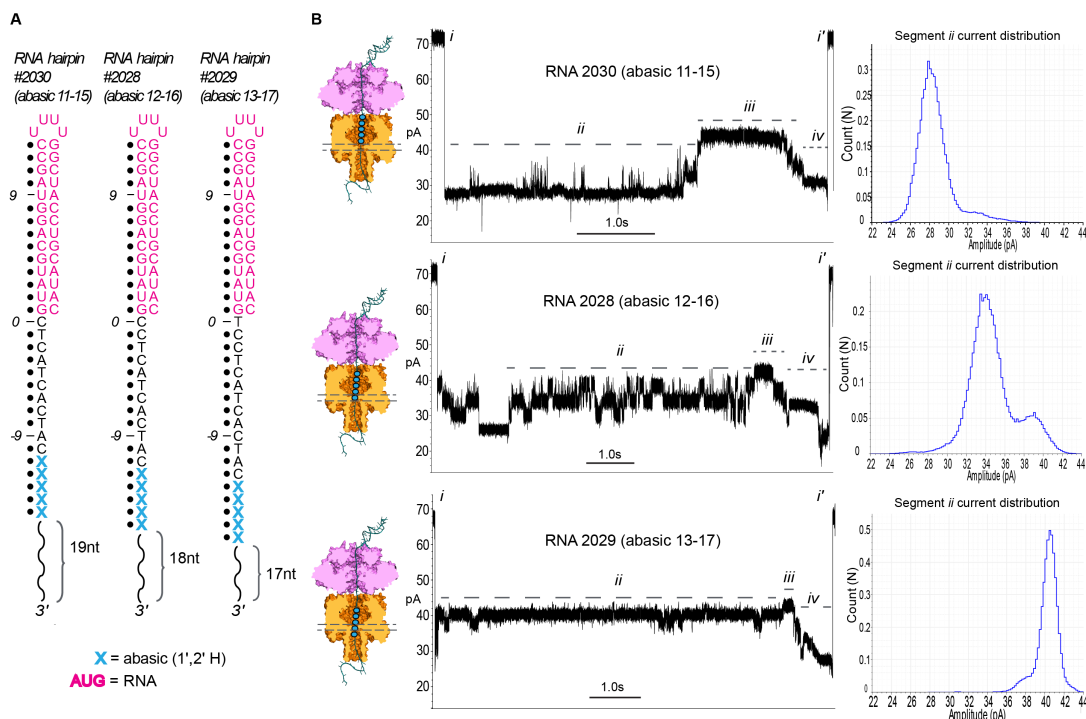


Figure 4.5: Amplitudes associated with single-nucleotide displacement of hairpin substrates can be identified from RNA induced-pausing

(A) RNA hairpin substrates designed to determine ionic current amplitudes associated with a single-nucleotide displacement. The position of the five abasic reporter is moved ± 1 nt relative to the RNA duplex in **Figure 4.4** (middle, #2028). This should position the reporter ± 1 nt within the aHL pore when LTag encounters the RNA duplex. To maintain the same overall length of each substrate, the 3' tail is shortened or extended by one nucleotide. Position of the five abasic block is indicated in parenthesis below each substrate and numbered relative to the first nucleotide of the RNA duplex. (B) Example traces for LTag helicase complexes with RNA hairpin substrates in (A). Steps in each trace are labeled as in **Figure 4.3**. The RNA-induced pausing of the LTag/hairpin complex for each substrate is labeled *ii*. A conceptual illustration for this state is shown to the left of each trace. The first of the abasic reporter residues, indicated as blue circles, start proximal to the sensitive portion of the aHL nanopore (indicated by horizontal dashed lines), and each substrate should advance the first abasic residue closer to the sensitive part of the nanopore. For each example trace, the ionic current amplitude distribution of step *ii* is shown to the right.

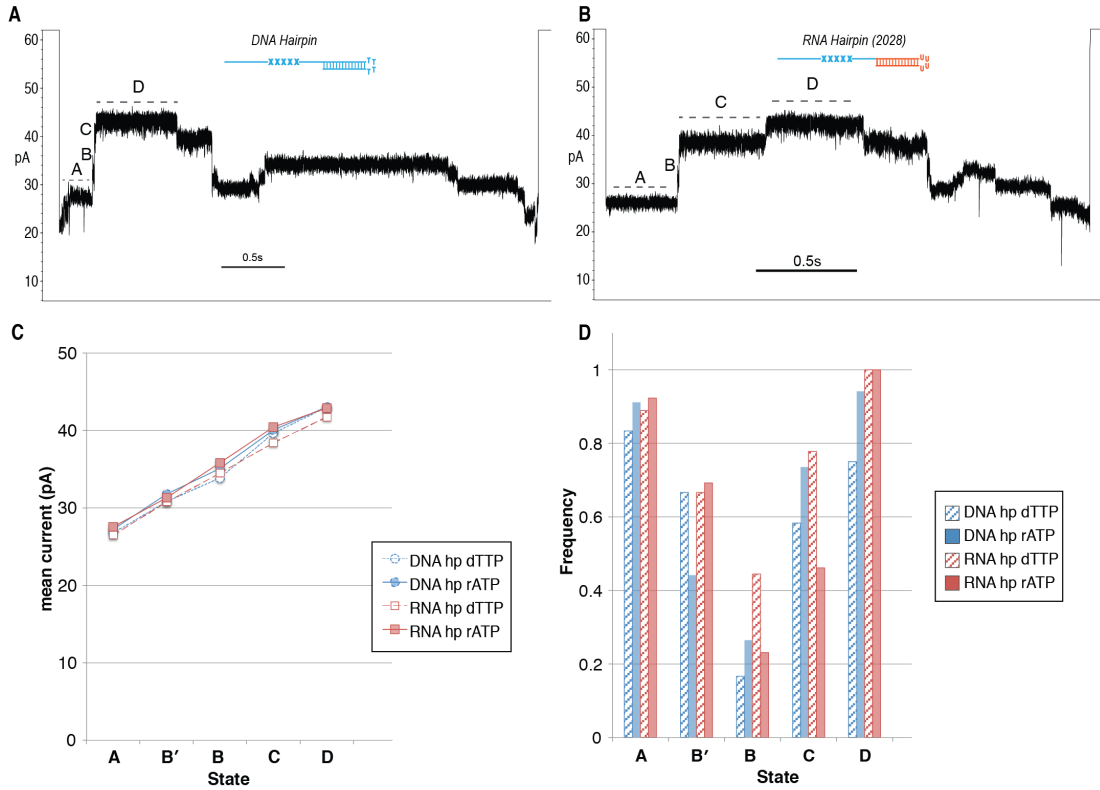


Figure 4.6: Reproducible ionic current amplitudes are associated with single-nucleotide displacement of DNA and RNA hairpin substrates during active LTag-mediated translocation through aHL nanopore

(A and B) show typical nanopore ionic current traces for LTag/hairpin complexes with hairpin #2025 (DNA) and #2028 (RNA/DNA chimera) during initial translocation of the abasic reporter through the sensitive part of aHL nanopore. This should correlate with helicase translocation through hairpin substrate duplex region. The lettered segments correspond amplitudes for 3 sequential nucleotide displacements of first abasic residue entering the sensitive portion of aHL nanopore as shown in **Figure 4.5** (A, 28pA; B, 34pA; C, 40pA) and the highest amplitude state (D, 42pA), which corresponds to multiple abasic residues in the nanopore. Schematic of the substrates are shown above both representative traces. A distinct alternative amplitude for state B observed in some events, called B' = 31pA, is not shown in either of the traces. (C) Median amplitudes measured from multiple enzyme/nucleic acid complexes. The median amplitude for each step as measured by hand-segmentation of events for either RNA or DNA hairpin substrate. The number of complexes measured for each substrate with either NTP present: #2025 substrate 34 (rATP) and 12 events (dTTP), #2028 substrate 13 (rATP) and 18 (dTTP). rATP or dTTP are present each of the experiments at 2mM. State B' was included to account for a distinct amplitude from amplitude B. Both states B and B' were observed in some events with RNA and DNA substrates. (D) indicates the normalized frequency that each state was observed in events analyzed for the state mean amplitudes shown in C.

both RNA and DNA substrates with either rATP or dTTP as an energy source for LTag. It is notable that an additional state needed to be added (referred to as B', ~31 pA) to account for an amplitude distinct from state B (~34 pA). This occurred between state A (~28 pA) and state C (~40 pA), and was not mutually exclusive with state B. The frequency that these states was observed in events displaying a high amplitude state was less than 1 for all states except state D.

Translocation rate of LTag on DNA substrates captured in a nanopore can be estimated from high amplitude markers

The previous experiment established a correlation between ionic current state, the position of the strand in the pore and the LTag helicase. With this in mind we sought to measure the translocation rate of LTag helicase when bound to strands captured in the pore. To do this we designed another DNA hairpin substrate that contained both the original 5-abasic block and a C9 PEG spacer, with the intention that this should produce two sequential high amplitude ionic current signatures. The C9 spacer was placed 13 nt away from the 3'-most abasic residue (**Figure 4.7A**). Blockade events like the one shown in **Figure 4.7B** were typical (~20% of all blockades >0.2 seconds). The second high amplitude was indeed produced downstream in these blockades as predicted. These ionic current features defined 13 nt passing through the aHL pore, and are presumed to correspond with the helicase moving 13nt towards the 5' end of the hairpin. To calculate LTag helicase translocation rate, duration between the high amplitude markers (indicated by arrows in **Figure 4.7B**) was measured and divided by the number of nucleotides between the two features (13 nt). Rates were calculated with varying concentrations of dTTP under otherwise similar conditions to all previous nanopore experiments with LTag

helicase. The calculated translocation rate was 3.3nt/second with 2mM dTTP (concentration typically used in most of the previous experiments) (**Figure 4.7C**). There was an increase in nt/sec translocation rate with increased concentration of dTTP concentration, as might be expected from an active helicase controlling movement of the DNA strand through the nanopore.

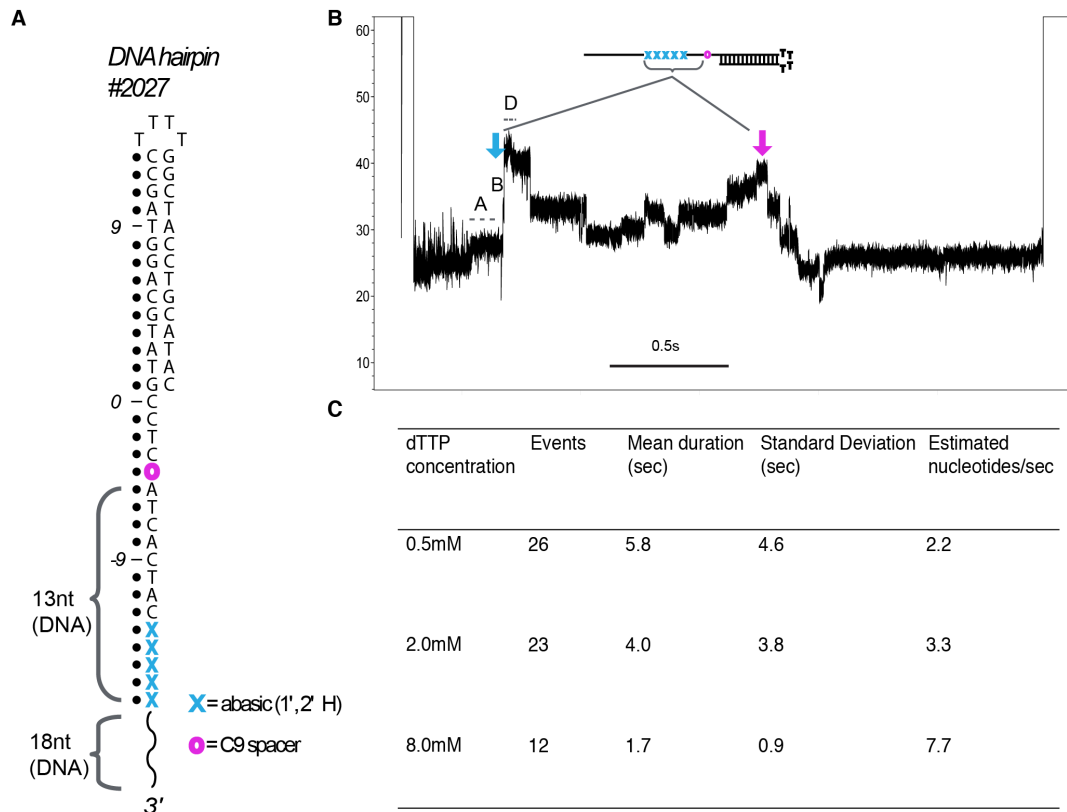


Figure 4.7: Approximate DNA translocation rates by LTag helicase fragment atop aHL nanopore measured over 13nt window defined by two high amplitude markers

(A) shows a schematic of the hairpin DNA substrate being analyzed in the single-channel nanopore experiments. The substrate is identical to DNA hairpin #2025 shown in **Figure 4.3**, with the exception of a C9 PEG spacer that is placed 13nt downstream from the first abasic residue that would enter the pore upon capture of the substrate 3' end. (B) shows a typical ionic current trace during capture of this substrate in the presence of active LTag helicase. The ionic current amplitude associated with entry of the first abasic in the sensitive part of aHL nanopore is indicated by the blue arrow. The magenta arrow indicates the second high amplitude marker from the C9 spacer passing through the aHL pore. These two amplitudes are associated with these two features in the #2027 substrate and define boundaries for 13nt of DNA passing through the nanopore. (C) shows estimates of strand

translocation rates with LTag bound under different NTP concentrations. The rates are calculated using the approximate 13nt window defined by the ionic signatures described in (B) and the duration between these two events passing through the nanopore. The mean nucleotide per second rate calculated from the number of events sampled is shown in the right most column.

Conclusion

In total, these results support the conclusion that LTag helicase fragment is capable of forming active complexes that can move DNA and RNA strands through a nanopore. The dwell time extension of LTag/DNA complexes in the presence of non-hydrolyzable nucleotide analog γ -S-ATP is consistent with this idea. Interestingly, strands containing only RNA were not readily used by LTag in our translocation assay. The series of RNA/DNA chimeric hairpin substrates indicate that while LTag helicase may translocate along RNA, there is an inhibitory effect on LTag helicase translocation. This would be consistent with published NTP hydrolysis measurements, which show that full-length LTag has reduced stimulation of NTPase activity with RNA relative to DNA (102).

The results of shifting the abasic reporter in the RNA hairpin substrates indicate that the LTag helicase construct used here may take single nucleotide steps atop aHL nanopore. The sequence of amplitudes very closely resembles the sequence for single nucleotide advancement of DNA through aHL pore with phi29 DNA polymerase (10). However, the low frequency of observing the all amplitudes associated with the single nucleotide increment movements suggests that LTag may not always advance in single-nt steps. The necessary addition of a B' state suggests sub-steps in the catalytic cycle might contribute to the observed ionic current pattern, as has been described for Hel308 helicase with MspA nanopore (12). Interestingly, the translocation rates estimated for LTag helicase atop aHL pore (3.3nt/second) are consistent with previously reported rates from single molecule unwinding reactions (116).

It must be noted that there is still uncertainty in the field as to the exact orientation and path of substrate DNA through LTag helicase domain during unwinding. Given that BPV E1 and LTag have recently been shown to unwind by strand exclusion (116, 117) and that E1 orientation of translocation is with the N-terminal domain advancing first, the models shown here for LTag helicase unwinding duplex atop the nanopore seem plausible. It is also possible that LTag helicase fragment used here could also not be functioning as an intact hexamer. Further work (such as creating a chained construct similar to a previously described ClpX construct) (118) would be necessary to show that complete hexamers are actively functioning on top of the pore.

Methods

Oligonucleotide synthesis and purification

The LTag 262-627 helicase domain construct was designed essentially as described (Gai et al, 2003). Expression and purification followed the protocol outlined in Appendix B.

Oligonucleotide synthesis and purification

All oligonucleotides were synthesized by the Stanford Protein and Nucleic Acids facility (PAN) using standard phosphoramidite chemistry. Oligonucleotides were purified by denaturing 7M urea polyacrylamide gel electrophoresis (PAGE) in 1x TBE, followed by overnight elution from an excised band at 4°C in 300mM Sodium Acetate pH 5.2 and 1mM EDTA. DNA was precipitated and recovered by adding 100% molecular biology grade ethanol (Sigma Aldrich) to 70% final v/v and centrifuged for 30 minutes at 14,000 x g and 4°C. Alternately, RNA-containing oligonucleotides were recovered by precipitation in 75% ethanol v/v and centrifuged at 14,000 x g for 30 minutes at 4°C. The ethanol mixture was aspirated and

oligonucleotides were then washed with an equal volume of 70% or 75% ethanol and pelleted again for 10 minutes at 14,000 x g and 4°C. Ethanol was aspirated and the pellets were allowed to dry under vacuum to remove residual ethanol. Oligonucleotides were then resuspended in nuclease free water, quantified by Nanodrop (Thermo Scientific), and stored at -80°C.

Hairpin nucleotide oligo preparation

Hairpin substrates were prepared at 10μM in 10mM Tris-HCl pH 8 and 50mM NaCl. The mixture was heated to 95°C for thirty seconds and snap cooled in an ice bath.

Nanopore experiments

Nanopore experiments were performed using a single α -HL nanopore (indicated by a step-wise increase in ionic current, range 68.0-72.5pA) embedded in a planar 1,2-diphytanoyl-syn-glycero-phosphatidylcholine bilayer established on a ~25μm aperture, as described previously (Akeson et al. 1999). Experiments were conducted in 0.3M KCl and 20mM HEPES pH 7.5 at 30°C (+/- 0.4°C) at 180mV (*trans* well positive). Dithiothreitol (2mM final) and MgCl₂ (10mM final) were added to the well on the *cis* side of the bilayer. Nucleic acid substrates were added to 2μM to the *cis* well. When NTPs were present, they were typically added to a final concentration of 2mM, unless otherwise noted. For experiments where LTag helicase was present, it was added to a final concentration of 240nM. Ionic current measurements were collected with an Axon Axopatch 200B (Molecular Devices) patch-clamp amplifier in whole-cell, voltage-clamped mode and filtered with an

analog low-pass Bessel filter at 5kHz, then digitized using an Axon Digidata 1440A analog-to-digital converter (Molecular Devices) at 100kHz bandwidth.

NTPase assays

NTPase assays used a malachite green colormetric assay essentially as described (119). Reactions of 250 μ l were prepared with 20mM HEPES pH 7.5 buffer solutions containing 2mM DTT, 1mM dTTP, 10mM MgCl₂, 30nM LTag hexamer, and varying amounts of KCl. When RNA was present a polyU 50mer was used at 1 μ M. Reactions were performed at temperatures indicated by incubating samples in a dry heat block. Reactions were initiated with the addition of the LTag hexamer. Individual time points were removed at 0, 8, 16, and 32 minutes. A standard curve was created from dilutions of K₂HPO₄ and used to calculate the amount of phosphate present in each reaction time point. Linear fits to the times points were used to calculate the rate of phosphate released by hydrolysis of dTTP.

Single-stranded translocation assay

Single-stranded translocation assays were performed similar to those described previously described for LTag helicase (111). Oligo nucleotides used were ordered from IDT and used the sequences: Oligo2; 5'~BiosG~UUUUUCCCCG**CAACC**AGATTGTGAGATCTGATTTTTTTTTTTTTTTT~3' and 5'~BiosG~TTTTTCCCCG**CAACC**AGATTGTGAGATCTGATTTTTTTTTTTTTTTTTT~3', where bolded nucleotides are RNA and BiosG is a 5' biotin. Reaction buffers contained with 20mM HEPES pH 7.5, 2mM DTT, 1mM dTTP, 10mM MgCl₂ and varying amounts of KCl. Displacement reactions were setup by incubating 2.0pmol

oligo with 38pmol streptavidin (NEB) in 18µl reaction buffer and heating to 45°C for 5 minutes. Excess biotin (900 pmol) in 1.5µl reaction buffer was added and the streptavidin-bound oligo mixture was allowed to sit at room temperature for 3 minutes. LTag hexamer (12pmol) was then added and the reaction was incubated at room temp for 10 minutes. The reaction was quenched with 40µl loading dye containing ficoll and 10mM EDTA. Bound and displace oligos were then analyzed on a 10% non-denaturing TBE acrylamide gel.

Acknowledgements

This work would not have been possible with out the help of Robin Abu-Shumays and Max Genetti.

Bibliography

1. Song,L., Hobaugh,M.R., Shustak,C., Cheley,S., Bayley,H. and Gouaux,J.E. (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science*, **274**, 1859–66.
2. Kasianowicz,J.J. and Bezrukov,S.M. (2016) On ‘three decades of nanopore sequencing’. *Nat. Biotechnol.*, **34**, 481–482.
3. Kasianowicz,J.J., Brandin,E., Branton,D. and Deamer,D.W. (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci.*, **93**, 13770–13773.
4. Wanunu,M., Dadosh,T., Ray,V., Jin,J., McReynolds,L. and Drndić,M. (2010) Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol.*, **5**, 807–814.
5. Akeson,M., Branton,D., Kasianowicz,J.J., Brandin,E. and Deamer,D.W. (1999) Microsecond Time-Scale Discrimination Among Polycytidylic Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments Within Single RNA Molecules. *Biophys. J.*, **77**, 3227–3233.
6. Branton,D., Deamer,D.W., Marziali,A., Bayley,H., Benner,S. a, Butler,T., Di Ventra,M., Garaj,S., Hibbs,A., Huang,X., *et al.* (2008) The potential and challenges of nanopore sequencing. *Nat. Biotechnol.*, **26**, 1146–53.
7. Benner,S., Chen,R.J.A., Wilson,N.A., Abu-Shumays,R., Hurt,N., Lieberman,K.R., Deamer,D.W., Dunbar,W.B. and Akeson,M. (2007) Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nanotechnol.*, **2**, 718–24.
8. Olasagasti,F., Lieberman,K.R., Benner,S., Cherf,G.M., Dahl,J.M., Deamer,D.W. and Akeson,M. (2010) Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.*, **5**, 798–806.

9. Lieberman,K.R., Cherf,G.M., Doody,M.J., Olasagasti,F., Kolodji,Y. and Akeson,M. (2010) Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J. Am. Chem. Soc.*, **132**, 17961–17972.
10. Cherf,G.M., Lieberman,K.R., Rashid,H., Lam,C.E., Karplus,K. and Akeson,M. (2012) Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.*, **30**, 344–348.
11. Manrao,E.A., Derrington,I.M., Laszlo,A.H., Langford,K.W., Hopper,M.K., Gillgren,N., Pavlenok,M., Niederweis,M. and Gundlach,J.H. (2012) Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.*, **30**, 349–353.
12. Derrington,I.M., Craig,J.M., Stava,E., Laszlo,A.H., Ross,B.C., Brinkerhoff,H., Nova,I.C., Doering,K., Tickman,B.I., Ronaghi,M., *et al.* (2015) Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.*, **33**, 1073–1075.
13. Jain,M., Olsen,H.E., Paten,B. and Akeson,M. (2016) Erratum to: The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.*, **17**, 256.
14. Jain,M., Koren,S., Quick,J., Rand,A.C., Sasani,T.A., Tyson,J.R., Beggs,A.D., Dilthey,A.T., Fiddes,I.T., Malla,S., *et al.* (2017) Nanopore sequencing and assembly of a human genome with ultra-long reads. *bioRxiv*, 10.1101/128835.
15. Ip,C.L.C., Loose,M., Tyson,J.R., de Cesare,M., Brown,B.L., Jain,M., Leggett,R.M., Eccles,D.A., Zalunin,V., Urban,J.M., *et al.* (2015) MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, 10.12688/f1000research.7201.1.
16. Butler,T.Z., Pavlenok,M., Derrington,I.M., Niederweis,M. and Gundlach,J.H. (2008) Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci.*, **105**, 20647–20652.

17. Goyal,P., Krasteva,P. V., Van Gerven,N., Gubellini,F., Van den Broeck,I., Troupiotis-Tsaïlaki,A., Jonckheere,W., Péhau-Arnaudet,G., Pinkner,J.S., Chapman,M.R., *et al.* (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature*, **516**, 250–253.
18. González-Pérez,A., Stibius,K.B., Vissing,T., Nielsen,C.H. and Mouritsen,O.G. (2009) Biomimetic triblock copolymer membrane arrays: A stable template for functional membrane proteins. *Langmuir*, **25**, 10447–10450.
19. Ayub,M. and Bayley,H. (2012) Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.*, **12**, 5637–43.
20. Ayub,M., Hardwick,S.W., Luisi,B.F. and Bayley,H. (2013) Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano Lett.*, **13**, 6144–50.
21. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
22. Nagalakshmi,U., Wang,Z., Waern,K., Shou,C., Raha,D., Gerstein,M. and Snyder,M. (2008) The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science (80-.)*, **320**, 1344–1349.
23. Marguerat,S. and Bähler,J. (2010) RNA-seq: From technology to biology. *Cell. Mol. Life Sci.*, **67**, 569–579.
24. Au,K.F., Sebastiano,V., Afshar,P.T., Durruthy,J.D., Lee,L., Williams,B.A., van Bakel,H., Schadt,E.E., Reijo-Pera,R.A., Underwood,J.G., *et al.* (2013) Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl. Acad. Sci.*, **110**, E4821–E4830.
25. Li,X., Xiong,X. and Yi,C. (2016) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods*, **14**, 23–31.
26. Behm-Ansmant,I., Helm,M. and Motorin,Y. (2011) Use of Specific Chemical Reagents for Detection of Modified Nucleotides in RNA. *J. Nucleic Acids*, **2011**, 1–17.

27. Lovejoy,A.F., Riordan,D.P. and Brown,P.O. (2014) Transcriptome-wide mapping of pseudouridines: Pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One*, **9**, e110799.
28. Dominissini,D., Moshitch-Moshkovitz,S., Schwartz,S., Salmon-Divon,M., Ungar,L., Osenberg,S., Cesarkas,K., Jacob-Hirsch,J., Amariglio,N., Kupiec,M., *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
29. Chen,K., Lu,Z., Wang,X., Fu,Y., Luo,G.Z., Liu,N., Han,D., Dominissini,D., Dai,Q., Pan,T., *et al.* (2015) High-resolution N6-methyladenosine (m6A) map using photo-crosslinking-assisted m6A sequencing. *Angew. Chemie - Int. Ed.*, **54**, 1587–1590.
30. Machnicka,M.A., Milanowska,K., Oglou,O.O., Purta,E., Kurkowska,M., Olchowik,A., Januszewski,W., Kalinowski,S., Dunin-Horkawicz,S., Rother,K.M., *et al.* (2013) MODOMICS: A database of RNA modification pathways - 2013 update. *Nucleic Acids Res.*, **41**, D262-7.
31. Sergeeva,O. V., Bogdanov,A.A. and Sergiev,P. V. (2015) What do we know about ribosomal RNA methylation in *Escherichia coli*? *Biochimie*, **117**, 110–8.
32. Motorin,Y. and Helm,M. (2010) tRNA Stabilization by Modified Nucleotides. *Biochemistry*, **49**, 4934–4944.
33. Suzuki,T., Nagao,A. and Suzuki,T. (2011) Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. *Annu. Rev. Genet.*, **45**, 299–329.
34. Kopajtich,R., Nicholls,T.J., Rorbach,J., Metodiev,M.D., Freisinger,P., Mandel,H., Vanlander,A., Ghezzi,D., Carrozzo,R., Taylor,R.W., *et al.* (2014) Mutations in GTPBP3 cause a mitochondrial translation defect associated with hypertrophic cardiomyopathy, lactic acidosis, and encephalopathy. *Am. J. Hum. Genet.*, **95**, 708–720.
35. Tischner,C., Hofer,A., Wulff,V., Stepek,J., Dumitru,I., Becker,L., Haack,T.,

- Kremer,L., Datta,A.N., Sperl,W., *et al.* (2015) MTO1 mediates tissue specificity of OXPHOS defects via tRNA modification and translation optimization, which can be bypassed by dietary intervention. *Hum. Mol. Genet.*, **24**, 2247–2266.
36. Rodriguez-Hernandez,A., Spears,J.L., Gaston,K.W., Limbach,P.A., Gamper,H., Hou,Y.M., Kaiser,R., Agris,P.F. and Perona,J.J. (2013) Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position. *J. Mol. Biol.*, **425**, 3888–3906.
 37. Kurata,S., Weixlbaumer,A., Ohtsuki,T., Shimazaki,T., Wada,T., Kirino,Y., Takai,K., Watanabe,K., Ramakrishnan,V. and Suzuki,T. (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J. Biol. Chem.*, **283**, 18801–11.
 38. Cantara,W.A., Murphy,F. V., Demirci,H. and Agris,P.F. (2013) Expanded use of sense codons is regulated by modified cytidines in tRNA. *Proc. Natl. Acad. Sci.*, **110**, 10964–10969.
 39. Suzuki,T. and Suzuki,T. (2014) A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic Acids Res.*, **42**, 7346–7357.
 40. Tsutomu,S., Asutaka,N. and Takeo,S. (2011) Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs. *Wiley Interdiscip. Rev. RNA*, **2**, 376–386.
 41. Kirino,Y., Yasukawa,T., Ohta,S., Akira,S., Ishihara,K., Watanabe,K. and Suzuki,T. (2004) Codon-specific translational defect caused by a wobble modification deficiency in mutant tRNA from a human mitochondrial disease. *Proc. Natl. Acad. Sci.*, **101**, 15070–15075.
 42. Van Haute,L., Dietmann,S., Kremer,L., Hussain,S., Pearce,S.F., Powell,C.A., Rorbach,J., Lantaff,R., Blanco,S., Sauer,S., *et al.* (2016) Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying

- mutations in NSUN3. *Nat. Commun.*, **7**, 12039.
43. Haag,S., Sloan,K.E., Ranjan,N., Warda,A.S., Kretschmer,J., Blessing,C., Hübner,B., Seikowski,J., Dennerlein,S., Rehling,P., *et al.* (2016) NSUN3 and ABH1 modify the wobble position of mt-tRNA^{Met} to expand codon recognition in mitochondrial translation. *EMBO J.*, **421**, e201694885.
 44. Yasukawa,T., Suzuki,T., Ishii,N., Ohta,S. and Watanabe,K. (2001) Wobble modification defect in tRNA disturbs codon-anticodon interaction in a mitochondrial disease. *EMBO J.*, **20**, 4794–802.
 45. Oszolak,F., Platt,A.R., Jones,D.R., Reifengerger,J.G., Sass,L.E., McInerney,P., Thompson,J.F., Bowers,J., Jarosz,M. and Milos,P.M. (2009) Direct RNA sequencing. *Nature*, **461**, 814–8.
 46. Vilfan,I.D., Tsai,Y.-C., Clark,T.A., Wegener,J., Dai,Q., Yi,C., Pan,T., Turner,S.W. and Korlach,J. (2013) Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J. Nanobiotechnology*, **11**, 8.
 47. Garalde,D.R., Snell,E.A., Jachimowicz,D., Heron,A.J., Bruce,M., Lloyd,J., Warland,A., Pantic,N., Admassu,T., Ciccone,J., *et al.* (2016) Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv*, 10.1101/068809.
 48. Rand,A.C., Jain,M., Eizenga,J.M., Musselman-Brown,A., Olsen,H.E., Akeson,M. and Paten,B. (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods*, **14**, 411–413.
 49. Phizicky,E.M. and Hopper,A.K. (2010) tRNA biology charges to the front. *Genes Dev.*, **24**, 1832–1860.
 50. Gustilo,E.M., Vendeix,F.A. and Agris,P.F. (2008) tRNA's modifications bring order to gene expression. *Curr. Opin. Microbiol.*, **11**, 134–140.
 51. Yi,C. and Pan,T. (2011) Cellular dynamics of RNA modification. *Acc. Chem. Res.*, **44**, 1380–1388.
 52. Raina,M. and Ibba,M. (2014) TRNAs as regulators of biological processes. *Front.*

Genet., **5**, 1–14.

53. Dittmar,K.A., Goodenbour,J.M. and Pan,T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, 2107–2115.
54. Chan,P.P., Cozen,A.E. and Lowe,T.M. (2011) Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol.*, **12**, R38.
55. Omer,A.D., Lowe,T.M., Russell,A.G., Ebhardt,H., Eddy,S.R. and Dennis,P.P. (2000) Homologs of Small Nucleolar RNAs in Archaea. *Science (80-.)*, **288**, 517–522.
56. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
57. Bernick,D.L., Dennis,P.P., Höchsmann,M. and Lowe,T.M. (2012) Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. *RNA*, **18**, 402–11.
58. Schaefer,M., Pollex,T., Hanna,K. and Lyko,F. (2009) RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.*, **37**, e12.
59. Hiley,S.L., Jackman,J., Babak,T., Trocheset,M., Morris,Q.D., Phizicky,E. and Hughes,T.R. (2005) Detection and discovery of RNA modifications using microarrays. *Nucleic Acids Res.*, **33**, e2.
60. Saikia,M., Fu,Y., Pavon-Eternod,M., He,C. and Pan,T. (2010) Genome-wide analysis of N1-methyl-adenosine modification in human tRNAs. *RNA*, **16**, 1317–27.
61. Chan,C.T.Y., Dyavaiah,M., DeMott,M.S., Taghizadeh,K., Dedon,P.C. and Begley,T.J. (2010) A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.*, **6**, 1–9.
62. Chan,C.T.Y., Pang,Y.L.J., Deng,W., Babu,I.R., Dyavaiah,M., Begley,T.J. and Dedon,P.C. (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.*,

3, 937.

63. Laszlo,A.H., Derrington,I.M., Brinkerhoff,H., Langford,K.W., Nova,I.C., Samson,J.M., Bartlett,J.J., Pavlenok,M. and Gundlach,J.H. (2013) Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci.*, **110**, 18904–18909.
64. Schreiber,J., Wescoe,Z.L., Abu-Shumays,R., Vivian,J.T., Baatar,B., Karplus,K. and Akeson,M. (2013) Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc. Natl. Acad. Sci.*, **110**, 18910–18915.
65. RajBhandary,U.L. (1994) Initiator transfer RNAs. [Review]. *J Bacteriol*, **176**, 547–552.
66. Deamer,D.W. and Branton,D. (2002) Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.*, **35**, 817–25.
67. Stein, a and Crothers,D.M. (1976) Conformational Changes of Transfer RNA. The Role of Magnesium(II). *Biochemistry*, **15**, 160–8.
68. Serebrov,V., Clarke,R.J., Gross,H.J. and Kisselev,L. (2001) Mg²⁺-induced tRNA folding. *Biochemistry*, **40**, 6688–98.
69. Gyarfas,B., Olasagasti,F., Benner,S., Garalde,D., Lieberman,K.R. and Akeson,M. (2009) Mapping the position of DNA polymerase-bound DNA templates in a nanopore at 5 ?? resolution. *ACS Nano*, **3**, 1457–1466.
70. Lieberman,K.R., Dahl,J.M., Mai,A.H., Akeson,M. and Wang,H. (2012) Dynamics of the translocation step measured in individual DNA polymerase complexes. *J. Am. Chem. Soc.*, **134**, 18816–18823.
71. Cortes,C. and Vapnik,V. (1995) Support-Vector Networks. *Mach. Learn.*, **20**, 273–297.
72. Laszlo,A.H., Derrington,I.M., Ross,B.C., Brinkerhoff,H., Adey,A., Nova,I.C., Craig,J.M., Langford,K.W., Samson,J.M., Daza,R., *et al.* (2014) Decoding long nanopore sequencing reads of natural DNA. *Nat. Biotechnol.*, **32**, 829–833.

73. Wescoe,Z.L., Schreiber,J. and Akeson,M. (2014) Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc.*, **136**, 16582–16587.
74. Karatzoglou,A., Smola,A., Hornik,K. and Zeileis,A. (2004) kernlab – An S4 Package for Kernel Methods in R. *J. Stat. Softw.*, **11**, 1–20.
75. Woese,C.R. and Fox,G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.*, **74**, 5088–5090.
76. Clarridge,J.E. and Alerts,C. (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.*, **17**, 840–862.
77. Smith,A.M., Abu-Shumays,R., Akeson,M. and Bernick,D.L. (2015) Capture, Unfolding, and Detection of Individual tRNA Molecules Using a Nanopore Device. *Front. Bioeng. Biotechnol.*, **3**, 1–11.
78. Ozsolak,F. and Milos,P.M. (2011) Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdiscip. Rev. RNA*, **2**, 565–570.
79. Moazed,D. and Noller,H.F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, **327**, 389–394.
80. Shine,J. and Dalgarno,L. (1974) The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci.*, **71**, 1342–1346.
81. Nakagawa,S., Niimura,Y., Miura,K. -i. and Gojobori,T. (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc. Natl. Acad. Sci.*, **107**, 6382–6387.
82. Jain,M., Fiddes,I.T., Miga,K.H., Olsen,H.E., Paten,B. and Akeson,M. (2015) Improved data analysis for the MinION nanopore sequencer. *Nat. Methods*, **12**, 351–356.
83. Cornelis,S., Gansemans,Y., Deleye,L., Deforce,D., Van Nieuwerburgh,F., Rand,A.C., Jain,M., Eizenga,J.M., Musselman-Brown,A., Olsen,H.E., *et al.* (2016) Scaffolding and Completing Genome Assemblies in Real-time with

- Nanopore Sequencing. *Neoplasia (United States)*, **8**, 1–6.
84. Moazed,D. and Noller,H.F. (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell*, **47**, 985–994.
 85. Zueva,V.S., Mankin,A.S., Bogdanov,A.A. and Baratova,L.A. (1985) Specific fragmentation of tRNA and rRNA at a 7-methylguanine residue in the presence of methylated carrier RNA. *Eur. J. Biochem.*, **146**, 679–87.
 86. Okamoto,S., Tamaru,A., Nakajima,C., Nishimura,K., Tanaka,Y., Tokuyama,S., Suzuki,Y. and Ochi,K. (2007) Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol. Microbiol.*, **63**, 1096–1106.
 87. Beauclerk,A.A.D. and Cundliffe,E. (1987) Sites of action of two ribosomal RNA methylases responsible for resistance to aminoglycosides. *J. Mol. Biol.*, **193**, 661–671.
 88. Doi,Y., Yokoyama,K., Yamane,K., Wachino,J., Shibata,N., Yagi,T., Shibayama,K., Kato,H. and Arakawa,Y. (2004) Plasmid-mediated 16S rRNA methylase in *Serratia marcescens* conferring high-level resistance to aminoglycosides. *Antimicrob. Agents Chemother.*, **48**, 491–6.
 89. Wachino,J.I. and Arakawa,Y. (2012) Exogenously acquired 16S rRNA methyltransferases found in aminoglycoside-resistant pathogenic Gram-negative bacteria: An update. *Drug Resist. Updat.*, **15**, 133–148.
 90. Bakin,A., Kowalak,J.A., Mccloskey,J.A. and Ofengand,J. (1994) The single pseudouridine residue in Escherichia coli 16S RNA is located at position 516. *Nucleic Acids Res.*, **22**, 3681–3684.
 91. Yarza,P., Yilmaz,P., Pruesse,E., Glöckner,F.O., Ludwig,W., Schleifer,K.-H., Whitman,W.B., Euzéby,J., Amann,R. and Rosselló-Móra,R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol. Microbiol.*, **12**, 635–45.
 92. Dennis,P.P. and Bremer,H. (2008) Modulation of Chemical Composition and

- Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, **3**.
93. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
 94. Loman,N.J. and Quinlan,A.R. (2014) Poretools: A toolkit for analyzing nanopore sequence data. *Bioinformatics*, **30**, 3399–3401.
 95. Li,H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN].
 96. Stoiber,M.H., Quick,J., Egan,R., Lee,J.E., Celniker,S.E., Neely,R., Loman,N., Pennacchio,L. and Brown,J.B. (2016) De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv*, **30**, 94672.
 97. Sović,I., Šikić,M., Wilm,A., Fenlon,S.N., Chen,S. and Nagarajan,N. (2016) Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat. Commun.*, **7**, 11307.
 98. Mastrangelo,I.A., Hough,P. V, Wall,J.S., Dodson,M., Dean,F.B. and Hurwitz,J. (1989) ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. *Nature*, **338**, 658–62.
 99. Wessel,R., Schweizer,J. and Stahl,H. (1992) Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J. Virol.*, **66**, 804–815.
 100. Li,D., Zhao,R., Lilyestrom,W., Gai,D., Zhang,R., DeCaprio,J. a, Fanning,E., Jochimiak,A., Szakonyi,G. and Chen,X.S. (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature*, **423**, 512–518.
 101. Stahl,H., Dröge,P. and Knippers,R. (1986) DNA helicase activity of SV40 large tumor antigen. *EMBO J.*, **5**, 1939–1944.

102. Scheffner,M., Knippers,R. and Stahl,H. (1989) RNA unwinding activity of SV40 large T antigen. *Cell*, **57**, 955–963.
103. Scheffner,M., Knippers,R. and Stahl,H. (1991) Simian-virus-40 large-T-antigen-catalyzed DNA and RNA unwinding reactions. *Eur. J. Biochem.*, **195**, 49–54.
104. Gai,D., Zhao,R., Li,D., Finkelstein,C. V. and Chen,X.S. (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell*, **119**, 47–60.
105. Mancini,E.J. and Tuma,R. (2012) Viral Molecular Machines. **726**.
106. Yu,X.J., Greenleaf,W.B., Shi,Y.S. and Chen,X.S. (2015) Mechanism of subunit coordination of an AAA+ hexameric molecular nanomachine. *Nanomedicine Nanotechnology, Biol. Med.*, **11**, 531–541.
107. Singleton,M.R., Sawaya,M.R., Ellenberger,T. and Wigley,D.B. (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, **101**, 589–600.
108. Enemark,E.J. and Joshua-Tor,L. (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, **442**, 270–275.
109. Gai,D., Li,D., Finkelstein,C. V., Ott,R.D., Taneja,P., Fanning,E. and Chen,X.S. (2004) Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J. Biol. Chem.*, **279**, 38952–38959.
110. Giacherio,D. and Hager,L.P. (1979) A poly(dT)-stimulated ATPase activity associated with simian virus 40 large T antigen. *J. Biol. Chem.*, **254**, 8113–8116.
111. Morris,P.D., Byrd,A.K., Tackett,A.J., Cameron,C.E., Tanega,P., Ott,R., Fanning,E. and Raney,K.D. (2002) Hepatitis C virus NS3 and simian virus 40 T antigen helicases displace streptavidin from 5'-biotinylated oligonucleotides but not from 3'-biotinylated oligonucleotides: Evidence for directional bias in translocation on single-stranded DNA. *Biochemistry*, **41**, 2372–2378.

112. Dahl,J.M., Mai,A.H., Cherf,G.M., Jetha,N.N., Garalde,D.R., Marziali,A., Akeson,M., Wang,H. and Lieberman,K.R. (2012) Direct observation of translocation in individual DNA polymerase complexes. *J. Biol. Chem.*, **287**, 13407–13421.
113. Dahl,J.M., Wang,H., Lázaro,J.M., Salas,M. and Lieberman,K.R. (2014) Dynamics of translocation and substrate binding in individual complexes formed with active site mutants of ϕ 29 DNA polymerase. *J. Biol. Chem.*, **289**, 6350–6361.
114. Stoddart,D., Heron,A.J., Klingelhoefer,J., Mikhailova,E., Maglia,G. and Bayley,H. (2010) Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano Lett.*, **10**, 3633–7.
115. Perera,R.T., Fleming,A.M., Peterson,A.M., Heemstra,J.M., Burrows,C.J. and White,H.S. (2016) Unzipping of A-Form DNA-RNA, A-Form DNA-PNA, and B-Form DNA-DNA in the alpha-Hemolysin Nanopore. *Biophys. J.*, **110**, 306–314.
116. Yardimci,H., Wang,X., Loveland,A.B., Tappin,I., Rudner,D.Z., Hurwitz,J., van Oijen,A.M. and Walter,J.C. (2012) Bypass of a protein barrier by a replicative DNA helicase. *Nature*, **492**, 205–209.
117. Lee,S.-J., Syed,S., Enemark,E.J., Schuck,S., Stenlund,A., Ha,T. and Joshua-Tor,L. (2014) Dynamic look at DNA unwinding by a replicative helicase. *Proc. Natl. Acad. Sci.*, **111**, E827–E835.
118. Martin,A., Baker,T.A. and Sauer,R.T. (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature*, **437**, 1115–1120.
119. Chan,K.M., Delfert,D. and Junger,K.D. (1986) A direct colorimetric assay for Ca^{2+} -stimulated ATPase activity. *Anal. Biochem.*, **157**, 375–80.

Appendix A

Protocol for selective preparation of RNA for nanopore sequencing

Version 1 (05/17/17)

The library preparation generally follows Oxford Nanopore Technologies protocol DRS_9026_v1_revA_15Dec2016 from community.nanoporetech.com

Required ONT components:

- Direct RNA Sequencing Kit (SQK-RNA001) (this kit contains a RTA primary adapter with 3' poly(dT) with a 5' desthiobiotin for use with polyA RNA. For other RNA classes a customized adapter needs to be designed in advance.)
- Flow Cells FLO-MIN106 (R9.4 version)
- MinION sequencer

The present ONT direct RNA sequencing method requires sequential ligation of two adapters to the 3' end of the RNA being targeted for sequencing. The first adapter is double stranded DNA and is referred to by ONT literature as the RT adapter (RTA). One strand of the adapter is complementary to the 3' end of the target RNA (eg. Oligo(dT) for polyA RNA) and acts as a splint during the first ligation reaction. This 1st ligation is necessary to covalently link the RNA 3' end to the other strand of the adapter (see ONT literature for further description). This adapter can be customized to target other RNA classes by changing the 3' end sequence to be complementary to the target RNA's 3' end. The splint strand can also serve as a primer for the optional reverse transcription step. The second adapter is provided by ONT in the kit and is

preloaded with the RNA processing motor protein. This adapter has a 3' overhang that is complementary to the 3' end of the RTA top strand and is not customizable.

1. RNA extraction (60 min)

Use TriZol or similar reagent according generally following the manufacturer recommended protocol.

Note: extending incubation time of Tri reagent with cells to 20 minutes during lysis has improved recovery of ribosomal RNAs, and may generally improve recovery of large RNA from RNPs. Also note, a secondary extraction of the aqueous phase with pure chloroform after initial phase separation step reduces phenol carryover.

After final ethanol wash resuspend extracted crude RNA in NSF water.

Quantitate RNA using Nanodrop or similar instrument

2. DNase treatment and phenol cleanup (90 min)

Note: Use RNase-free DNase I, like NEB cat. #M0303. DNase I Buffer: 10mM TrisHCl pH 7.5, 0.5mM EDTA, 2.5mM MgCl₂, 1.0mM CaCl₂ may or may not be supplied. Prepare a 10x buffer in advance if not supplied with the enzyme.

Using up to 100ug crude RNA, add to 1x DNase I buffer in 150ul.

Add 2U DNaseI/10ug crude RNA

Incubate 37°C x 10 min

End reaction by adding 1.5ul of 0.5 M EDTA (final concentration of 5 mM)

Clean up reaction with an equal volume of 1:1 phenol/chloroform (pH 8).

Incubate 5 min at RT and then separate phases by centrifugation

Retain aqueous phase and repeat extraction with pure chloroform. Phase separate and retain aqueous phase

Ethanol precipitate with 80% EtOH and NaOAc. Chill -80°C for 20 min to precipitate. Spin 40 min @ 4°C to pellet. Wash 1x with 80% EtOH. Aspirate and air dry pellet. Resuspend Total RNA pellet in NSF water.

3. RNA selection by hybridization of RTA to target RNA (60 min)

Note: This step will selectively hybridize the primary (RT) adapter to the target RNA a mixture. The adapter desthiobiotin is bound on magnetic beads conjugated to streptavidin. This allows background RNA to be washed away from target RNA. Target RNA is then eluted from beads with the adapter and carried forward in subsequent library preparations. This step takes the place of

other purification techniques. Buffer recipes for bead binding and washing are provided below and will need to be prepared in advance.

Buffers:

- 10X TNE: 100mM Tris-HCl pH 8, 10mM EDTA, 500mM NaCl. (commercial 10x TE can be used for TE componets)
- Buffer A: 100mM NaOH, 50mM NaCl
- Buffer B: 100mM NaCl
- Buffer HSBW: 1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA, 0.025% v/v NP-40
- Buffer BW2: 100mM NaCl, 10mM Tris-HCl pH8, 1mM EDTA
- Buffer EB: 100mM NaCl, 10mM Tris-HCl pH8, 1mM EDTA, 0.05mM Biotin

Hybridization reaction (recommended to place all components into 200 µl PCR tube)

Reagent	of [Stock]	Add	to [Final]
RNA		___ µl	Est. 1pmol target RNA 3' ends (for 16S rRNA in total RNA ~ 5000ng total RNA)
Primary adapter	10µM	1.5 µl	15 pmol
TNE buffer (1X: TE buffer, 50mM NaCl)	10x	1.0 µl	1x
Nuclease-free water		___ µl	
TOTAL		10.0 µl	

- Incubate at 50°C for 1 minute, then slow cool to 23°C (room temp) (~10 mins in thermocycler at 5% ramp).

While hybridization reaction is going prepare 50 µl MyOne C1 (1250pmol free biotin binding capacity) streptavidin beads for use In 1.5ml Protein Lobind microfuge tube (Eppendorf # 022431081):

- Collect beads until solution clears on magnet rack (ThermoFisher Cat# 12321D or similar) and discard supernatant.
- Wash beads with 2x vol. Buffer A by resuspending the beads. Collect beads on rack and discard buffer. Repeat 1x.
- Wash beads with 2x vol. Buffer B by resuspending the beads. Collect beads on rack and discard buffer. Repeat 1x.
- Wash beads with 2x vol. Buffer HSBW by resuspending the beads. Collect beads on rack and discard buffer.
- Resuspend beads in 1x vol (50 µl) Buffer HSBW. Add 40ul NSF water to beads and mix for beads in 0.5X HSBW

Add hybridization reaction to previously prepared beads, mix well by pipet and allow adapter to bind beads by incubating 20 minutes at room temp.

- After incubation, collect beads on magnet and remove supernatant.

Resuspend beads 100 µl Buffer HSBW to wash and remove background RNA.

- Repeat bead collection step as above.

Resuspend beads 100 µl Buffer BW2 to wash and remove background RNA.

- Repeat collection step as above.

To elute target RNA and adapter, resuspend beads in 10 µl Buffer EB. Incubate 30°C for 20 minutes, then collect beads on magnet and pipet supernatant into clean 1.5ml Lo Bind Eppendorf tube. The supernatant contains target RNA and adapter, which is carried forward into the primary ligation step.

4. Primary adapter ligation (15 min):

Reagent	of [Stock]	Add	to [Final]
RNA hybridization		10 µl	
5x Quick Ligation Reaction Buffer (NEB)	5x	3.0 µl	1x
Nuclease-free water		0.5 µl	
T4 DNA Ligase (NEB)	2000 U/µl	1.5 µl	3000 U

TOTAL		15.0 μ l	
-------	--	--------------	--

- Incubate at room temperature for 10 min.
- After incubation, transfer all 15 μ l of the RT Adapter ligation mixture into reverse transcription mixture or if skipping RT step move on to clean-up described at the end of

5. Reverse Transcription (optional, 90min or 20min skipping RT):

Note: This step is optional. If skipping, bring Primary ligation (Step 5) to 40 μ l with ~25 μ l NSF and follow clean-up described at the end of this step. If performing reverse transcription, add TGIRT transcriptase last after mixing the other reagents.

RT reaction in 200 μ l PCR tube:

Reagent	of [Stock]	Add	to [Final]
“Primary adapter ligation” mixture from step 4		15 μ l	
5x TGIRT 300mM KCl Buffer	5x	8 μ l	1x
DTT	0.1 M	4 μ l	10 mM
dNTPs	10 mM each	2 μ l	0.5 mM each
Nuclease-free water		9 μ l	
TGIRT		2 μ l	
TOTAL		40 μ l	

- Incubate at 41°C for 50 min in thermocycler, then add 0.5 μ l 0.5M EDTA and heat 70°C for 5 min. Bring back down to 4°C before SPRI purification.

Reaction clean-up step:

- add 1.8x vol Agencourt RNAClean XP SPRI beads to reaction, incubate 5min on beads, then separate on magnet, and remove supernatant.
- Wash once using 150 μ l of 70% ethanol in nuclease-free water by rotating the tube twice by 180° and waiting for the sample to collect toward magnet between rotations), and remove supernatant.
- Elute into 20 μ l of NSF water, holding 5 minutes before collecting beads on magnet and transferring sample to clean 1.5ml Lo-Bind tube

6. Sequencing adapter ligation (30 min):

Reagent	of [Stock]	Add	to [Final]
RT RNA from step 2		20 µl	
RMX sequencing adapter		6 µl	
5x Quick Ligation Reaction Buffer (NEB)	5x	8 µl	1x
Nuclease-free water		3 µl	
T4 DNAL (NEB)	2000 U/ul	3 µl	150 U/ul
TOTAL		40 µl	

- Incubate at room temperature for 10 min.

Reaction clean-up:

- 1.0x Agencourt RNAClean XP purify to ligation reaction
- Wash twice with 150 µl supplied ONT SPRI wash buffer, rotate 180° x 2 on magnet. Remove supernatant.
- Elute by resuspending beads in 21µl of ONT SPRI elution buffer, holding 10 minutes before collecting beads on magnet and transferring supernatant to clean 1.5ml Lo Bind tube.

This is your final library

7. MinION (15 min):

Note: RBF1 (Running Buffer Fuel Mix 1) is supplied as a 2x concentrate.

Note: This protocol uses SpotON loading. Follow flowcell preparation outlined at [Community.nanoporetech.com](https://community.nanoporetech.com)

Load ~150 ng of library per flowcell.

Prepare 1.1ml 1x RBF1 (550ul 2x RBF1 and 550ul NSF water)

Flush flowcell with 0.8ml 1x RBF1 through priming/load port, close port when done

Library preparation for single loading (75ul):

- Add 37.5 μ l of 2x RBF1 to volume of necessary for 150ng total sample
- Add nuclease-free water (non-DEPC treated) to bring final volume to 75 μ l

At least 10 min after 0.8ml flush, open SpotOn port and add 200ul 1x RBF1 through priming/loading port. Leave port open.

To load, add 75ul library drop-wise to SpotOn port (capillary action should pull drops into flowcell), close SpotOn port, then close priming/loading port

Initiate run from MinKNOW software on PC that the MinION is connected to.

Appendix B

Protocol for preparation of LTag 262-627 helicase fragment

LT 262-627 helicase expression and purification protocol

By Andrew Smith and Max Genneti

This is a purification protocol developed for LT helicase 262-627 with an N-terminal GST-6His tag. The GST-LT fusion protein was expressed in BL21 E. coli. The full purification from start to finish takes four days. In this protocol cell culture is done at a scale of 2L. Typical yield of purified protein at the end of the size exclusion is ~0.1-0.2 mg/L culture. This may be improved by other lysis methods. This protocol is based on the protocol for expression of LT 262-627 published by Gai et al. 2004.

Buffers:

Lysis Buffer (100ml):

250 mM NaCl
50 mM Tris.Cl
1 mM DTT
1 mM PMSF
4 units/ml DNase I (NEB M0303S)
pH 8

Wash Buffer (300ml) :

250 mM NaCl
50 mM Tris.Cl
1 mM DTT
pH 8

Elution Buffer (15ml):

250 mM NaCl
50 mM Tris.Cl
10 mM Reduced Glutathione (0.046g powdered stock)
1 mM DTT
pH 8

FPLC Buffer:

250mM NaCl
50mM Tris-HCl pH 8
1mM DTT

A filtered 500ml stock of Wash Buffer (no DTT) should be prepared the night of the 2nd day. Smaller volumes for the other buffers can be taken from this stock and amended with appropriate amounts of other reagents to make the rest of the buffers mentioned here.

Day 1: Overnight culture

1. Overnight culture:
 - a. Start an overnight culture of *E.coli* BL21 pAS6.0 (GST-LT) in 25 ml of Lysogeny Broth + 50 µg/ml Kanamycin (LB/Kan) in a 50 ml Falcon.
 - b. Grow culture overnight at 37° C in shaking incubator at 280 RPM.

Current cell line used for this is dated 8/28/14 and stored in -80C freezer box #1 “cell lines”, made by Max Genetti. Carefully scrape from frozen cells with a sterile pipet tip and place into media.

If using Falcon tubes as suggested here, cap should be loosely capped and secured with tape. Tightly capped cultures will become anoxic and fail to grow.

Day 2: Large scale culture and expression

2. Inoculate 4 x 500 ml (2L total) LB/Kan in a 1L Erlenmeyer flask with 1% inoculum (5ml) from starter culture. Grow at 37°C, shaking 280 RPM to an OD₆₀₀ of ~0.6 (about 3 hours).

Flasks of LB media should be pre-warmed in incubator ~1 hr. prior to inoculation to ensure exponential growth.

Note: cultures growth will vary. Cultures grown in baffled flasks have grown to appropriate density in ~2hrs.

3. Induction of protein expression.
 - a. For each flask, induce expression with 500 µl of 1 M IPTG, final concentration of 1 mM IPTG.
 - b. Reduce the temperature of the incubator to 25° C.
 - c. Grow overnight for 18hr, 280 RPM.

At the end of the day, wash buffer (500mL) without DTT should be prepared, filtered, and placed in the fridge. This can be used the next day

upon the addition of appropriate amounts of concentrated DTT (1M). This pre-prepared buffer will also be used in the next day's work to make 15ml of Elution buffer by adding solid reduced glutathione to 10mM and in subsequent column purification steps on Day 4.

Day 3: Lysis and Affinity purification

4. Harvest cells by centrifugation at 3,000 g for 30 min at 4° C

After the 18 hour expression, transfer the cultures to 1 L centrifuge containers designed for use with JLA 9.100 centrifuge rotor or other similar rotor. Balance bottles accordingly using a pan balance. Spin in Beckman Avanti J series centrifuge or similar.

5. Resuspend cell pellets in 30ml cold Lysis Buffer per 1L culture volume and transfer to 50ml Falcon tubes.

PMSF and DNaseI should be added to the buffer just after resuspending cells. Do not add these reagents if you are planning to freeze and store the cells.

The volume of buffer can be varied depending on the size of the cell pellets.

Note: The prep can be suspended long-term at this point. Flash freeze the resuspended cells in liquid nitrogen and store at -80°C.

6. Lyse cells using a sonicator.

Lyse in 50ml falcon 30sec on, 59sec off, 50% power x 5. Make sure to check tube for damage from sonication, which can occur if sonicator probe contacts the tube. If damage is suspected do not use tube in following centrifugation step, but instead transfer lysate to new tube

Alternately, 2-3 passes through a pressure-based lysis method like a cell disruptor or French press can be used, which has increased final yield of purified protein 4-5 fold.

7. Clarify the lysate by centrifugation 14000 g for 60 min at 4° C.
 - a. Lysate should be in 50ml falcon tubes rated for >14000 g and balanced between tubes
 - b. Use Fiberlite F250 rotor in Beckman Avanti J Series centrifuge or similar.

After this step lysate can be filtered with a 0.2µm PES filter to ensure that all cells have been removed, or alternately, immediately transfer clarified lysate into fresh tubes by pipet, being very careful avoid the cell debris at the bottom of the centrifuge tube.

8. Prepare Glutathione Resin (performed on ice or at 4° C)
 - a. Transfer 3ml suspended reduced glutathione sepharose resin (GE Healthcare 17-0756-01) into 15ml Falcon tube and equilibrate with 5ml Wash Buffer.
 - b. Spin down resin 500 g x 2 in Eppendorf table top centrifuge or similar
 - c. Remove supernatant
 - d. Repeat wash step above (a-c) x 3
9. Affinity purify GST-LT fusion protein on prepared glutathione resin (all steps performed on ice or at 4° C):
 - a. Incubate lysate and resin 30-60 min at 4°C on a gentle rocker.
 - b. Load the incubated resin onto a 2 ml Pierce Spin Column (or similar) with 30 µm pores

It is recommended to do this step using gentle centrifugation (~20 g) in a table top centrifuge, but it can be don by gravity flow, which will increase the amount of time it takes to the lysate/resin, which is typically >60ml at this point

- c. Wash the column with 10 column volumes (c.v.) (15ml) Wash buffer, using gravity flow.
- d. Elute bound GST-LT protein with 0.5 c.v. x 8 (750µl x 8) Elution Buffer

The elution step is done in small increments to keep the protein as concentrated as possible. The first 750µl should be added and then centrifuged gently to collect it quickly. It will not likely contain much protein. Retain this in a catch tube, quickly stopper the column, add the next 0.5 c.v. Elution Buffer and incubate on ice for 15min. Continue eluting the protein in 0.5 c.v. increments saving each individually, checking for protein by Nanodrop or some other method. Typically most of the protein is removed from the column by fraction 8 (>0.1mg/ml concentration), but large batches of protein yield more protein with additional elutions. Keep all elution fractions on ice! These should be analyzed by SDS-PAGE at the end of the day.

- e. Incubate glutathione resin with 1 c.v. Elution Buffer to remove remaining bound protein, and wash thoroughly with Wash Buffer (~10

c.v.). Add additional Wash Buffer to cover , cap and store in the Pierce column at 4°C for use on the Day 4.

10. TEV Digest to remove GST tag from GST-LT fusion

- a. Combined elution fractions that contain protein in concentrations >0.1mg/ml in 15ml Falcon tube
- b. Add TurboTEV protease (Eton Biosciences 1500020012) to 1:50 w/w protease-to-GST-LT
- c. Incubate 10°C overnight in water bath found in BE 202 cold room.

Day 4: Reverse Affinity Purification and Size Exclusion purification of LT hexamers

11. Concentrate/Buffer Exchange TEV digest of GST-LT to remove reduced glutathione (Performed on ice or at 4° C):

- a. Prepare a 15 ml 10kDa MWCO spin concentrator (Millipore Amicon 15 or similar) by washing the membrane with MilliQ H₂O, and then spin 2-3ml wash buffer until mostly through the concentrator
- b. Transfer the digest to the concentrator and spin 3200 g until the volume is reduced to 1/10th original volume (typically 500-1000µl).

Retain flow through in reservoir for analysis by nanodrop or SDS-PAGE in case of protein loss.

- c. Dilute buffer by adding fresh wash buffer to spin concentrator to original volume of digest (5-10ml typically), pipetting gently to mix.
- d. Spin again 3200 g until the volume is reduced to again to 1/10th original volume
- e. Dilute buffer again as in step c., and spin until volume is reduced to again to volume to load on to Pierce column.

Incomplete removal of reduced glutathione during this step will result in inefficient removal of the cleaved GST tag in second purification step.

12. Equilibrate Reduced Glutathione Resin (Performed on ice or at 4° C):

Reuse the previous day's resin in the same column.

- a. Take 3 ml of suspended Reduced Glutathione Resin. Equilibrate with Wash Buffer in a 15 ml falcon.
- b. Add 5 ml Fusion Wash Buffer
- c. Centrifuge at 500 g for 3 min.

- d. Remove supernatant
 - e. Should result in ~1.5ml packed and equilibrated resin
13. Reverse affinity purification to remove GST tag and TEV protease (performed on ice or at 4° C):
- a. Load the incubated resin onto a 2 ml Spin Chromatography Column with 30 µm pores. Collect the Flow Through.
 - b. Wash the column with 4 volumes of 1.5 ml (4 X 1 column volumes) Wash Buffer.
 - c. Collect the Wash Fractions.
 - d. Elute with 4 volumes of 1.5 ml (4 X 1 column volumes) Elution Buffer.
 - e. Collect the first 1.5 ml fraction.
 - f. Stop the flow and incubate the column in the Elution Buffer for 15 minutes
 - g. Collect fraction 2-4.

The LT protein should be in the flow through and wash. GST tag and TEV protease should be in the elution fractions. Retain all fractions for later gel analysis.

14. Quantification:
- a. Perform SDS-PAGE on the Flow Through, Washes, and Elutions.
 - b. LT runs near 30 kDa.
 - c. Combine the fractions which contain LT.

15. Assemble LT hexamers:
- a. To the combined fractions containing LT, add ATP to 1 mM final concentration.
 - b. Incubate at 37° C for 30 minutes.
 - c. Concentrate down to less than 500 µl using a 30kDa spin column.

16. Size exclusion chromatography to isolate LT hexamers:

If this is the first time using FPLC please have an experienced user accompany you during this step. The FPLC and its columns are expensive and can be easily damage if care is not taken. A number of useful programs are already stored under user name AndrewS. Details of these programs are indicated in their name.

- a. Equilibrate Superdex 200 10/300 column on FPLC (GE Healthcare 17-5175-01) with Wash Buffer running program SD200_1.2cv_0.4ml/min

The column is stored long-term in 20% ETOH, which needs to be removed by an additional wash of MilliQ H2O prior to the buffer equilibration step.

- b. Wash 1ml loop with 3-4ml Wash Buffer by setting the 7-way valve to “Load” position and injecting from a syringe with a special blunt FPLC needle.
- c. Spin concentrated LT hexamers in 1.5ml microcentrifuge tube 10k g x 10min at 4°C to remove an possible precipitates
- d. Load LT hexamers on 1ml loop by syringe through the 7-way valve
- e. Run program SD200_1.1c.v_0.4ml/min_frcCollect

This program will collect 1ml fractions throughout the entire run. Hexamers typically elute from the column in fractions 12-13, after aggregates (fraction 9) and before the monomers (fractions 15-16).

- f. Quantitate protein in fractions 9-17 by Nanodrop or other method, and then analyze these fractions by SDS-PAGE, which should confirm the presence of LT protein in fractions 12-13
- g. Add 80% glycerol to a final concentration of 5% v/v final and concentrate in a 30kD MWCO spin concentrator. Ideally, concentrate to 10.7mg/ml for 24μM hexamer.

I have successfully concentrated hexamers to ~10mg/ml, and literature suggests that concentrations of 20mg/ml are possible. That said I have had the protein precipitate at 10mg/ml upon the addition of glycerol at 15% v/v final, which leads me to suggest adding the glycerol at lower concentration and prior to concentration.

- h. Measure A₂₈₀ by nanodrop, calculate, and record final molar concentration of hexamers
- i. Aliquot in volumes (<5μl) in 1.5ml Protein Low-bind Eppendorf tubes and flash freeze in liquid nitrogen. Store -80°C